SCUOLA INTERNAZIONALE
SUPERIORE di STUDI AVANZATI
International School
for Advanced Studies

Sounds in noise: Behavioral and neural studies of illusory continuity and discontinuity

Ekaterina Vinnik

PhD candidate

Advisor

Evan Balaban

Trieste, 30 September 2009

Моему дедушке

Acknowledgements

This work would not have been possible without the help and encouragement of my advisor Prof. Evan Balaban, whom I would like to thank for his immense professionalism and kindness.

I am grateful to my colleagues at the Cognitive Neuroscience sector of SISSA, especially to Prof. Mathew Diamond for inspiring discussions and constant support, Nicola van Rijsbergen, Damir Kovačić, Linda Sebastianutto, Valeria Caruso and Mahan Azadpour for all their help.

Numerous participants from Trieste University contributed their time (hours of psychophysical and electrophysiological experiments!) to this work, demonstrating inspiring enthusiasm for scientific research.

I would like to express my gratitude to the Cognitive Neuroscience Sector for the award of a SISSA fellowship, which gave me the possibility to fulfill this work. I have furthermore to thank the organizers and tutors of the Okinawa Computational Neuroscience Course 2009, as some of the ideas presented in the thesis were implemented during that course.

Last but not the least, thanks to Pavel Itskov for always being himself. Workwise, he is invariably the first one to debate new ideas and contributed greatly to this work. I'd like to thank my mother, whose love and personal example always encourages me.

Table of contents

Summary

The ability to parse an auditory scene into meaningful components varies greatly between individuals; some are able to parse out and write down competing musical pieces while others struggle to understand each word whenever they have to converse in a noisy environment. Using a simple discrimination task, healthy, normally-hearing adult participants were asked to judge whether a pure tone (with or without amplitude modulation) was continuous or contained a gap. One quarter of the participants consistently heard a gap when none was present, if the tone was accompanied by a higher-frequency noise burst with a lower edge beginning one octave away from the tone (that did not have any energy overlapping the tone). This novel form of informational masking (perceptual interference between components with non-overlapping sound energy) was named 'illusory auditory discontinuity'.

The phenomenon appears to reflect natural differences in auditory processing rather than differences in decision-making strategies because: (1) susceptibility to illusory discontinuity correlates with individual differences in auditory streaming (measured using a classical ABA sequential paradigm); and (2) electroencephalographic responses elicited by tones overlaid by short noise bursts (when these sounds are not the focus of attention) are significantly correlated with the occurrence of illusory auditory discontinuity in both an early event-related potential (ERP) component (40-66 ms), and a later ERP component (270-350 ms) after noise onset. Participants prone to illusory discontinuity also tended not to perceive the 'auditory continuity illusion' (in which a tone is heard continuing under a burst of noise centered on the tone frequency that completely masks it) at short noise durations, but reliably perceived the auditory continuity illusion at longer noise durations. These results suggest that a number of attributes describing how individuals differentially parse complex auditory scenes are related to individual differences in two potentially independent attributes of neural processing, reflected here by EEG waveform differences at ~50 msec and ~300 msec after noise onset.

Neural correlates of the auditory continuity illusion were also investigated by adjusting masker loudness, so that when listeners were given physically identical stimuli, they correctly detected the gap in a target tone on some trials, while on other trials they reported the tone as continuous (experiencing illusory continuity). Higher power of low-frequency EEG activity (in the delta-theta range, <6 Hz) was observed prior to the onset of tones that were subsequently judged as discontinuous, with no other consistent EEG differences found after the onset of tones. These data suggest that the occurrence of the continuity illusion may depend on the brain state that exists immediately before a trial begins.

INTRODUCTION

The ability to parse an auditory scene into meaningful components varies greatly between individuals; some are able to parse out and write down competing musical pieces while others struggle to understand each word whenever they have to converse in a noisy environment. Systematic physiological correlations with this substantial individual variation in extracting sounds from noise in the normal population have not yet been uncovered, and the mechanisms that may contribute to such variation remain to be elucidated.

In clinical populations several aspects of auditory feature encoding have been linked to difficulties in speech-in-noise hearing performance, with about 70% of the variance explained by hearing loss, spectral resolution differences, temporal resolution differences and the ability to recognize speech in auditory and visual maskers (Zekveld et al, 2007). However, the above-mentioned factors do not necessarily cover all performance aspects relevant for speech perception in noise (Houtgast and Festen, 2008); cognitive factors such as short term memory and general cognitive ability are also thought to play a role (reviewed in Akeroyd, 2008, Rajan and Cainer, 2008).

Since acoustic landscapes are complex and rapidly changing, one crucial property of the auditory system is its rapid and plastic ability for sound source segregation, predictive coding and detection of novel objects in complex scenes. Sensitivity to changes in the auditory scene, and the ability to extract meaningful information crucially depend on the detailed representation of previously existing auditory objects. If the signal is weak, it might be totally obscured by the components of the background in such a way that no signal detection strategy could recover the information that has been lost at the level of the auditory periphery. However, even large differences in spectral content and loudness of the components of an auditory scene do not guarantee perfect performance, and a mixture of dissimilar sounds can still be difficult to perceive and parse. This type of masking is thought to arise from an inability to correctly allocate the sound components to different sound streams based on recent experience and prior knowledge (Neff and Dethlefs, 1995).

It is possible that people who are less susceptible to the disruptive effects of noise are able to segregate sound sources better, and thus are better at selective attention to target sounds. An ***auditory streaming hypothesis*** for individual differences in auditory perceptual performance in the presence of noise would predict that differences between individuals may be based on differences in the context-dependent function of higher-order processes used in auditory scene analysis, as distinct from processes used in basic sensory feature encoding *per se*. The work reported here examined individual susceptibility to effects of noise in the context of an auditory stream segregation phenomenon called the auditory continuity illusion. The first two chapters of this thesis provide a review of the subject matter from the various areas of auditory and neural science that were incorporated into the design of the experiments described in Chapters 5, 6, and 7. The ideas behind auditory stream segregation are also reviewed, as well as proposals for the mechanisms by which the brain accomplishes streaming. The third chapter deals more particularly with the auditory continuity illusion, and the fourth chapter explores some ideas about unifying neural mechanisms that may play an important role in illusory continuity perception.

CHAPTER 1. Perception of sounds in noise and auditory streaming

Factors correlated with individual variation in speech in noise performance

The case of speech in noise is the best-studied example of noise effects on sound perception, and several factors have been reliably shown to affect performance, such as the ability of listeners to represent a sound in the absence of a masker, as well as specific processes that underlie context-driven plasticity in sound representation.

For example, children with reading impairments have been found to have less-detailed subcortical temporal encoding of formant transitions (Hornickel et al, 2009), and to be impaired in speech perception in noise (Bradlow et al, 2003). Hearing impaired listeners also show less precise temporal encoding that affects speech recognition in noise (Dreschler and Plomp, 1985, George et al, 2006). Spectral resolution in terms of effective bandwidth or upward spread of masking appears to be a less-important factor for both hearing impaired listeners (see Houtgast and Festen, 2008, for a review) and for the normally-hearing population (Oxenham et al, 2003).

Improved speech-in-noise perception in a population of healthy, normally-hearing individuals has been associated with higher levels of bilateral activation in Heschl's gyrus and lateral superior temporal gyrus (Binder, 2004). Speech restoration, an important component of speech-in-noise perception, is dependent on many aspects of auditory coding, one of which is the function of the efferent system. Participants with poor restoration of time-reversed speech showed reduced medial olivocochlear bundle (MOCB) asymmetry and functionality compared with highly performing individuals (Grataloup et al, 2009, de Boer and Thornton, 2008).

Informational masking

In most real-world environments, behaviorally important sounds are mixed with unwanted sounds, or 'maskers'. If the signal is weak and the masker is so loud that it completely obscures the target tone, the peripheral neural representation of the sounds is dominated by the masker to such a degree that it is not possible to reliably extract the signal. This extensively-studied type of masking can be called 'energetic masking', to contrast it with other types of signal-masker interaction which are not due to peripheral signal-to-noise constraints, but rather to an inability to correctly determine that the signal and the masker belong to separate auditory objects. This latter form of masking is referred to in the literature as 'informational masking' (IM, Watson et al, 1976, Pollack, 1975, Lutfi, 1989, Leek et al 1991, Kidd et al, 1994).

Two major forms of informational masking are distinguished – informational masking due to masker uncertainty, and due to target-masker similarity. Initially the term informational masking was proposed for an elevation in signal threshold resulting from masker novelty or uncertainty

(Watson et al, 1976). In these experiments, the maskers that were most poorly predicted from recent input statistics caused larger threshold elevations compared to the conditions in which the same maskers were drawn from a limited set of possible options or were the same on all trials. The association of IM with the effects of stimulus uncertainty forms the basis for most of the theoretical work on IM (Lutfi, 1993; Oh and Lutfi, 1998; Richards and Neff, 2004).

Another factor that influences informational masking is target-masker similarity in any perceptual dimension such as timbre, pitch or patterns of amplitude changes (Kidd et al, 1998, Kidd et al, 2002). Most studies on informational masking have measured the detection of the presence of a tone (e.g. Neff and Green, 1987, Leek and Watson, 1984, Neff, 1995, Oh and Lutfi, 1998) or discrimination of a difference in the frequency or intensity of a tone (e.g. Watson et al, 1976, Howard et al, 1984, Kidd et al, 1994, Leek et al, 1991, Neff and Jesteadt, 1996); speech discrimination has also been used (Brungart, 2001, Freyman et al, 2001, Wightman et al, 2005, Mattys et al, 2009). It has been suggested that the use of non-speech sounds for studies on informational masking is advantageous because this simplifies the handling of experimental variables, and allows investigators to place masker energy at frequencies remote from the signal (Kidd et al, 2001). In a typical paradigm the possibility of energetic masking is minimized by a 'protected region' surrounding the signal, in which a masker is never present. The informational masker usually consists of random-frequency chords (Neff et al, 1987, Lufti et al, 2003) or noise bursts (Bonino et al, 2008).

Susceptibility to informational masking is known to vary greatly in the population of healthy adults with normal hearing. Thresholds for detecting a pure tone differ by as much as 59 dB across listeners if random-frequency chords are presented simultaneously with the target (Neff and Dethlefs, 1995). Participants with the best performance show masking thresholds that are assumed to be determined primarily by the frequency selectivity of the peripheral auditory system (Neff and Dethlefs, 1995), in agreement with traditional models of masking (Fletcher, 1940, Leek et al 1991, Kidd et al, 1994, Patterson et al, 1982). It has been shown that the differences between individuals are not likely to be due to the width of the peripheral filters per se, since they are similar between participants that show diverse amounts of informational masking (Oxenham et al, 2003). Instead, it has been suggested that listeners susceptible to informational masking are more prone to the effects of masker uncertainty or target-masker similarity because of some perceptual factors, supposedly the reduced ability to independently process signals in different frequency regions. This is summed up in the psychoacoustic literature as a distinction between 'analytic listening' as opposed to 'holistic', or 'synthetic' listening (Neff and Dethlefs, 1995, Oxenham, 2003). Holistic listeners (the ones who show the largest amounts of informational masking) are consistent in their performance across experimental sessions and are sometimes resistant to attempts to improve their performance with training (Neff and Dethlefs, 1995). 'Holistic' and 'analytic' are convenient labels; however it is important to note that performance in detecting a target in a random-frequency multicomponent masker is a continuum rather than a bimodal distribution (Neff and Dethlefs, 1995, Lutfi et el, 2003).

Interestingly, a similar distinction between "holistic" and "analytical" listening exists in pitch perception. Holistic listening favours grouping a set of harmonically-related tones into a single pitch percept with a single fundamental frequency, whereas in "analytical" listening the harmonics can be “heard out” separately. If a set of three very short harmonically-related tones is played in rapid succession with no overlap, they will be heard “analytically” when presented in silence, and “holistically” (as a single continuous tone with a single fundamental frequency pitch percept) when presented in noise (Hall and Peters, 1981).

Lufti and coauthors (2003) proposed that under conditions of uncertainty listeners take the output of neighboring auditory filters into greater account (called ‘off-frequency listening’). The number and frequency range of additional filters that are used are presumed to vary from one individual to another, with a larger number of “off-frequency filters” adding additional variance to the energy detection process. While this idea may help to account for intersubject variability in task performance, no studies to date have examined the relationship between individual susceptibility to different forms of informational masking and other psychophysical and neural phenomena in healthy, normally-hearing individuals. In general, the physiological factors that underlie such differences in performance are currently unknown, but will hopefully be increasingly discovered as the physiological bases of stream segregation and cross-frequency integration are elucidated.

Age effects on susceptibility to informational masking

Children are thought to be more susceptible to informational masking. While well-trained adult listeners can show informational masking effects up to 59 dB (Neff and Dethlefs, 1995, Neff and Callaghan, 1987, 1988, Neff and Green, 1987, Oh and Lutfi 1998), effects in preschool-age children can exceed 70 dB (Oh et al, 2001). As is found with other auditory abilities such as frequency discrimination (Allen et al, 1989) and spectral pattern recognition (Allen and Wightman, 1992), the performance of late school-aged children is much closer to that of adults.

Quantitative, rather than qualitative differences in informational masking abilities may underlie age-related variability in performance. Lufti et al (2003) analyzed the performance of adults and children under conditions of informational masking using random-frequency multicomponent maskers. They concluded from principal components analysis that large individual differences at all ages reflect the influence of one single factor, which they suggested is the number of frequency filters taken into account by decision-making under conditions of uncertainty.

Estimating the effect of aging on the susceptibility to informational masking is not straightforward, due to the confounding effects of hearing loss and possible anatomical changes at the auditory periphery. Thus, the major reason for difficulties in understanding speech in noisy environments is age-dependent hearing loss (Houtgast and Festen, 2008), but as shown recently in individuals without hearing loss, cognitive factors also appear to play a role (Rajan and Cainer, 2008). Temporal resolution measures can also be a notable predictor of speech comprehension in noise, even when presentation level is accounted for. Interestingly, measures of spectral resolution are not strong predictors of speech comprehension abilities (see Houtgast and Festen (2008) for a review).

Studies have provided evidence for normal stream segregation of pure tones differing in frequency in healthy aging populations using both subjective (Snyder and Alain, 2007) and objective (Alain et al, 1996, Trainor and Trehub 1989) behavioral measures of streaming, in addition to ERP measures (Snyder and Alain, 2007). This suggests that mild peripheral hearing loss typical of aging does not impair streaming. Young adults perceived more streaming than older adults, but only for tones with peripherally resolved harmonics (Grimault et al, 2001), consistent with peripheral hearing deficits in the older adults. As suggested by Snyder and Alain (2007), speech perception in noise might be related to difficulties in the segregation of simultaneous, rather than sequentially-presented sounds.

CHAPTER 2. Neural bases of streaming

Basic attributes of streaming and mechanistic explanations for them

A fundamental problem faced by all auditory systems is the segregation of sound sources and their grouping into coherent auditory objects with behavioral meaning. Natural acoustic scenes tend to be composed of multiple auditory objects which may change over time, so both concurrent sound segregation and sequential streaming are essential.

Bregman (1990) proposed a distinction between primitive and schema-based processes in auditory scene analysis. The former were envisioned as stimulus-driven preattentive auditory processes that parse the incoming sound into acoustic elements or features, and group them into larger streams that are likely to originate from the same location and the same sound source. Schema-based processes, on the other hand, refer to perceptual grouping which is based on previous experience. This review first concentrates on the 'primitive', stimulus-driven mechanisms and then discusses some recent studies that attempt to determine the form in which this activity is "read out" to become the conscious sensations that we know from our everyday experience.

The earliest explanations of streaming in the frequency domain (van Noorden, 1975) stated that stream segregation is experienced when sounds excite distinct cochlear 'filters' or peripheral tonotopic 'channels'. Hartman and Johnson (1991) proposed that stream segregation depends on the degree of peripheral tonotopic separation but does not require a compete lack of the overlap. This line of thought led to the development of a number of computational models of auditory stream analysis (Beauvois and Meddis, 1996, Denham and Winkler, 2006) that are largely compatible with results from human psychophysical studies on streaming. The core of these models is the idea that stream segregation results from decorrelation of neuronal activity caused by the sounds that excite different populations of neurons at different times. Though initially this 'grouping by coactivation' theory was developed to account only for tonotopically separated (i.e., spatially segregated) inputs, it could be generalized to predict that separate streams will be perceived whenever sounds evoke uncorrelated responses along the dimensions represented in the auditory system, such as pitch and timbre (Cusack and Roberts, 1999, 2000, 2004, Roberts, Glasberg, Moore, 2002, Singh and Bregman, 1997, Vliegen and Oxenham, 1999), loudness (M.M.Rose and Moore, 2000), temporal modulation (Grimault et al, 2002, Gutschalk et al, 2007, Grimault et al., 2002) and perhaps also frequency modulation (but see Darwin, 2005).

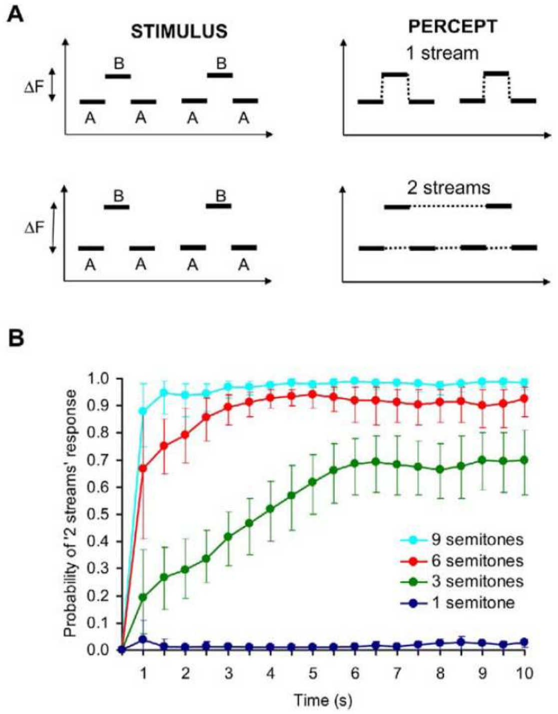


Figure 2.1 Auditory streaming (reproduced from Micheyl et al, 2005, with permission). **(A)** Stimulus sequences typically used to measure perceptual auditory streaming (left) and corresponding auditory percepts (right). **(B)** The build-up of auditory stream segregation. The different curves represent the proportion of "two streams" responses measured psychophysically in human subjects listening to sound sequences with different frequency separations (ΔFs) between the A and B tones.

The simplest way the segregation of sound sources without the use of sound localization cues can be modeled in the laboratory is the presentation of a sequence of fast alternating tones that differ in frequency. Depending on the frequency separation between the tones, presentation rate (duration of the tones and interstimulus intervals), as well as an individual's intentions (van Noorden, 1975), this sequence can be heard as a single stream of alternating tones or as two separate sound streams. Numerous psychophysical studies starting from Miller and Heise (1950) utilized these kinds of simple stimuli to characterize the dynamics of formation of auditory streams.

Four basic observations have been made in these studies that characterize the percepts and provide benchmarks that any neural account of streaming must explain.

(1) First, the greater the frequency separation between tones, the more probable it is that the sequence will be perceived as two streams.

(2) Second, a faster presentation rate and shorter duration of the tones facilitate perceptual segregation (Bregman et al, 2000).
(3) Third, a single stream is more often heard in the beginning of the sequence, with the probability of segregation increasing with the time of presentation, a phenomenon termed 'the build-up' of streaming (see Figure 2.1 B).
(4) Fourth, if a sequence with an intermediate frequency separation is presented for a long time, a listener's percept can switch spontaneously from that of a single stream to separate streams and back. Such shifts occur rapidly, and closely resemble visual bi-stable perceptual effects such as the Necker cube or Rubin vase.

Physiological mechanisms of streaming have been studied with the use of animal electrophysiology, human brain recordings and functional imaging. We start from animal recording experiments that have elucidated basic properties of the auditory sensory system that are crucial for sequential sound perception.

Fishman and colleagues (2001) recorded responses in primary auditory cortex of awake macaque monkeys to alternating low and high tones in an ABAB… pattern. The results indicated that the neurons with receptive fields that best matched the frequency of type A tones could respond to tones of frequency B presented in isolation, but responded less well to them in the ABAB sequence. Consistent with the psychophysical results, the suppression of the responses to B tones increased with increasing frequency separation between A and B, and was stronger when the sequence was presented at a fast tempo (more than 10 Hz). They concluded that neuronal stream segregation arises from temporal interactions between excitatory and subsequent inhibitory phases of the sensory response, with key features being the differential suppression following best-frequency (BF) and non-BF tones, and an increased susceptibility of response suppression of non-BF tones by preceding best-frequency tones as the non-BF tones get further away from a neuron's BF.

Similar results have been found in forward masking experiments, in which suppression of the responses to a repeated target stimulus was stronger if a preferred stimulus preceded the target (Brosch and Schneiner, 1997, Calford and Semple, 1995). Subsequent studies have confirmed the generality of these findings and extended them to non-primate species such as starlings (*Sturnus vulgaris*), who possess behavioral capabilities of auditory scene analysis similar to humans. Starlings show a similar pattern of neuronal responses in their primary auditory forebrain field L (specifically, in the input layer of the field L complex, L2), (Bee and Klump, 2004). A subsequent study by Micheyl and colleagues (2005) showed that neuronal activity recorded from macaque primary auditory cortex in response to ABA… sequences could quantitatively account for human psychophysical data, especially with regard to the increase in probability of segregation with increasing frequency separation and the build up of streaming over time. For the latter they examined the dynamics of neuronal responses to the stimuli as they unfolded in time. Their data analyses indicated that the perceptual build-up of streaming in humans correlated with the temporal dynamics of adaptation of the firing rate of macaque A1 neurons over seconds, which had a different time course for preferred (BF) and non-preferred (non-BF) stimuli. So, the main correlates of streaming in primary auditory cortical neurons appear to be based on frequency selectivity, differential suppression of neuronal responses by preferred and non-preferred stimuli, and multisecond adaptation that further weakens responses to non-preferred stimuli.

The detailed mechanistic origins of all three of these phenomena remain unclear. It has been suggested that circuit interactions involving local GABAergic inhibition may possess the right time constants to explain two-tone masking interactions (Brosch and Schreiner, 1997, Calford and Semple 1995, but see Wehr and Zador, 2005). Afterhyperpolarization due to inactivation of voltage-gated potassium channels (Locke and Nerbonne, 1997) may also provide a possible mechanism for cortical forward suppression. This effect decays over hundreds of milliseconds, although it can possibly last for tens of seconds (when potassium currents are activated by sodium ions (Schwindt, et al. 1989)). Wehr and Zador (2005) directly tested the inhibitory conductances hypothesis and found that inhibitory conductances rarely last longer than 50-100 ms, whereas spike responses can remain suppressed for several hundred milliseconds (under anesthesia that does not prolong the opening time of GABA–receptor coupled channels). They concluded that the time constants of GABA modulation of cellular firing rate are not consistent with the time course of two-tone suppression.

Observed decreases of firing rate with repeated stimulation have been attributed to other mechanisms, for example fast synaptic depression (possibly at the thalamocortical level) that would apply to all the synaptic inputs to the cell (Wehr and Zador, 2005). Alternatively the repetition suppression effect may be directly inherited from thalamic input (Denham, 2001, Carandini and Ferster, 1997). Studies from other sensory modalities have suggested that stimulus-specific adaptation might be an intrinsic property of single neurons rather than a circuit property, and may be due to calcium-dependent slow afterhyperpolarizations (Dıaz-Quesada and Maravall, 2008). The mechanisms underlying multisecond adaptation in the frequency-response functions of auditory cortex neurons also remain unknown (Ulanovsky et al, 2004).

The idea that forward suppression and multisecond adaptation in thalamo-cortical circuits are particularly crucial for auditory streaming has been challenged by recent evidence coming from recordings in guinea pigs. Pressnitzer et al (2008) showed that neurons in the cochlear nucleus, which receives direct projections from the cochlea, also exhibit forward suppression and multisecond adaptation even in anaesthetized animals. These recordings were done in the ventral part of the cochlear nucleus for which efferent projections are sparse (Winer, 2006), so the authors felt it is was unlikely that the suppression seen in the cochlear nucleus was a result of a top-down modulation. However, they did not rule out top-down effects mediated via the efferent olivocochlear sytem that would also change the responses of ventral cochlear nucleus neurons. The authors proposed that multisecond adaptation to tone sequences may result from some kind of interaction between neural processes with two different time constants: a synaptic depression mechanism with a long time constant, and a mechanism for fast recovery during the silent intervals.

In the context of a peripheral origin for and/or peripheral participation in streaming, it is important to note that forward masking (Relkin and Turner, 1998, Turner, 1994) and multisecond adaptation (Javel 1996) exist even at the level of auditory nerve fibers, although the relationship of these phenomena to streaming perception remains to be examined. Neural response adaptation was used in one of the earliest computational models of streaming (Beauvois and Meddis, 1996), which showed that frequency-based segregation in the auditory periphery could be used to model the build-up of streaming. Adaptation in the cochlear nucleus is thought to begin with the dynamics of neurotransmitter release at the synapse between a hair cell and the corresponding auditory nerve fiber (Zhang and Carney, 2005). As progressively more complex features are extracted along the auditory pathway, the adaptation of populations of cells that deal with more specialized features that emerge as a result of hierarchical processing could contribute to streaming based on adaptation to other sound properties beyond frequency.

Repetition suppression has been repeatedly shown to accompany formation of short- term memories about a stimulus in many sensory modalities (reviewed in Grill-Spector et al, 2005). By sharpening tuning curves and thus creating sparser representations, repetition suppression effects allow more efficient feature coding (Maravall et al, 2007) and encoding of temporal aspects of the stimulus. One interesting possibility is that repetition suppression effects may be accompanied by an increase in spike timing precision and inter-neuronal oscillatory coherence (Stopfer and Laurent, 1999), which could facilitate feature binding within population representations of single sensory objects and affect segregation of several simultaneously-presented but independent objects.

In summary, current experimental data provide evidence about how streams may be represented at the levels of neuronal populations at and below the primary auditory cortex. What remains unknown is the way in which this activity is utilized to form the conscious sensations that we know from our everyday experience. In the simplest case of sequential streaming it is possible to utilize response magnitude thresholds for the activity of single neurons to hypothesize how streaming might work. As shown by Micheyl et al (2005), the output of this simple model applied to the data from macaque primary auditory cortex showed good correspondence with the time course of the perceptual buildup of streaming in humans. However, realistic auditory scenes are much more complex, and different streams are formed by sounds that differ not only in pitch but in many other attributes, such as timbre and loudness. Elhilali and coauthors (2009) suggested the existence of neural mechanisms which could act as coincidence detectors over a broad range of sound feature representations. Their study did not find any cells in the primary auditory cortex of awake ferrets whose output was strongly influenced by temporal coherence across distant frequencies, but they suggested that such neurons may exist in other areas, or the coherence across channels might be determined by other network mechanisms (Elhilali et al, 2009). Advances in signal processing may boost our understanding of complex auditory scene analysis by providing predictions about the neural implementation of diverse algorithms that can later be tested.

Neural correlates of streaming in humans

Electrophysiological measures in humans have provided a very flexible tool to study streaming, as observed physiological changes can be directly related to perceptual reports (extensively reviewed in Micheyl et al, 2008). However, due to the poor spatial resolution of current techniques with respect to the tonotopic organization of the human auditory system, such studies do not allow investigators to isolate frequency-specific representations. This review will therefore be confined to studies that analyzed responses to physically identical stimuli perceived in different ways depending on the duration of the presentation (buildup of streaming) or stochastic changes in performance (the 'bistability' of streaming), as well as frequency separation-dependent changes in responses. (If one only considers frequency-dependent changes, it is problematic to disentangle variation in the stimuli from variation in the percept).

Gutschalk et al (2005) used magnetoencephalography (MEG) to examine how onset responses to sound change depending on whether a sound belongs to the same stream as a preceding sound or is a part of a different stream. They varied frequency separation between tones in an ABA-type sequence. If the frequency separation was small, B tones elicited smaller amplitudes of early components of evoked responses (called P_{1m} and N_{1m}) compared to the case when the frequency separation was larger. These results are consistent with the adaptation hypothesis, according to which tones B evoke smaller responses if they are preceded by A tones. This finding has been confirmed by another experiment in which the evoked magnetic fields were registered simultaneously with perceptual judgments while the frequency separation between tones A and B was kept constant. Spontaneous switches of perception from one stream to two streams covaried with the amplitude of the P_{1m} and N_{1m} components (Gutschalk et al, 2005). In this study, P1 peaked at 69-74 ms after the tone onset and N1 peaked at 109-115 msec after the tone onset. These midlatency components of auditory responses are thought to originate primarily from auditory cortices and associative cortices (lateral Heschl's gyrus, planum temporale and the superior temporal gyrus, Liegeois-Chauvel et al., 1994) Thus, a relationship between cortical activity during tone-triplet sequences and the perceptual organization of these sequences has also been found in human subjects.

Other studies that examined the effects of context on sound representation also stress the importance of early cortical evoked responses. The EEG P1 component peaks at approximately 50 ms after tone onset (from 35 to 75 msec) and is known to be suppressed in response to the repetition of a tone. P50 is present only during wakefulness and rapid eye movement (REM) sleep, and its amplitude is decreased by scopolamine, in contrast to the earlier Pa potential (15-40 ms poststimulus), thought to be localized to the primary auditory cortex, which does not habituate at high stimulation rates and is unaffected by scopolamine (Erwin and Buchwald, 1986). Human intracranial recordings have suggested that P50 arises from the auditory cortex (Ligeois-Chauvel, et al, 1994) and temporal cortices, mainly superior temporal gyrus (STG, Pelizzone et al, 1987, Reite et al, 1988), as well as frontal cortex (Weisser et al, 2001, Garcia-Rill et al, 2008). According to Huang and coauthors (2003), superior temporal gyrus (STG) sources account for 97% of scalp-recorded variance of the EEG vertex (Cz) signal in the 30-100 ms latency window in normal subjects. The amount of suppression depends on the presentation rate: the largest suppression is observed when the interstimulus interval is small, with suppression being abolished if the interstimulus interval is longer than 7 seconds. Similar phenomena are seen in both intracranial and scalp recordings (Rosburg et al, 2004). With an increasing number of presentations, the amplitude of the evoked potential is also affected (Haenschel et al, 2005). If a deviant is presented among the tones, the amplitude of P1 recovers (Boutros et al, 1995, Boutros et al, 1999).

The latency of P1 is much shorter that that of it's more famous counterpart, a derived component called mismatch negativity (MMN), which is elicited anywhere between 100-250 ms after the onset of the deviant stimulus (Garrido et al, 2009). MMN is elicited with a gap of up to 10 seconds between the standard tone and a deviant tone (Botther-Gander, Ulsperger 1992) – a somewhat longer time scale compared to P1. Several studies reported that the amplitude of MMN and P1 are correlated (e.g. Ermutly et al, 2007). Some authors have proposed that MMN indexes the contents of sensory memory, and P1 suppression might reflect adaptation-mediated saliency through cholinergic modulation of the sensory input, thus being important for awareness about the stimulus (see Reese et al, 1995 for a review).

Individuals with narcolepsy, a disease in which disturbances in arousal and attention are most obvious, have smaller P1 amplitudes in response to auditory stimulation, compared to normal controls (Boop at al, 2004). Both the amplitude of P50 and, more importantly, the degree of repetition suppression vary greatly in normal populations (Heinreichs, 2004, Anokhin et al, 2007). To date, the physiological mechanisms that underlie variation in the normal population in these parameters have not been systematically investigated. It is known, however, that in certain pathological conditions, such as schizophrenia and major depression (Sánchez-Morla et al, 2008), severe deficits of P1 suppression have been observed. In schizophrenia the degree of this functional deficit has been shown to covary with severity of the disorder (Baldeweg et al., 2004). Interestingly, basic auditory abilities such as streaming and perceptual restoration of missing sounds have also been demonstrated to be altered in schizophrenia (Nielzén and Olsson, 1997). In the absence of any direct experimental evidence linking the degree of P1 suppression deficits and alterations of auditory scene analysis, it makes sense to ask whether a tight correlation between the two might be present.

Streaming of concurrent sounds

All findings reviewed above consider sequential streaming, i.e. segregating components of auditory scene based on recent dynamics of the input. Another important aspect of auditory scene analysis is segregation of sounds presented simultaneously. Although the auditory system can benefit from predictions built upon preceding sounds, and use onset asynchrony to better parse auditory scenes, allocation of simultaneously presented features to their sources on the basis of spectral regularities (Roberts and Brunstrom, 1998) and modulation patterns (Moore, 1990) may also be beneficial.

In terms of sound segregation in modern urban environments, mechanical and multispeaker noise presents big challenges. Performance in a majority of auditory perceptual tasks drops under even moderate levels of noise in individuals with normal hearing (reviewed in Berglund et al, 1996, Henderson and Salvi, 1998), whereas in subjects with compromised auditory system function the effects of noise are highly detrimental (Larsby et al, 2005).

Interaction between perception of the target sounds and the noises are likely to involve a variety of feedforward and feedback processing mechanisms. It has been hypothesized that background noise may affect amplitude and temporal precision of the responses to target sounds (Galambos and Makeig, 1992, Alain et al, 2009). Surprisingly, noise can both enhance and inhibit neuronal responses, depending on the sound level: Galambos and Makeig (1992) have shown that contralateral noise reduced amplitudes of auditory responses in an intensity-dependent manner, whereas ipsilateral noise enhanced the amplitudes at low levels and depressed them at high levels. Moderate levels of ipsilateral noise also reduced the latency of auditory steady state responses (aSSR), a neurophysiological EEG or MEG response recorded to repeated clicks or amplitude modulated sounds.

Coherent amplitude modulations can improve signal detection in noise, a phenomenon termed ‘comodulation masking release’ (CMR). One neural mechanism proposed for this improvement involves wide-band inhibition driven by the noise itself. Thought to originate from the cells in the auditory periphery that receive broad band input (Pressnitzer et al., 2001, Winter et al., 2005), such inhibition enhances target detection by phase-locked suppression of responses to noise. Successive refinement and improvement of CMR-related properties has also been observed in primary auditory cortex neurons (Las et al., 2005)

Natural sounds tend to be complex, often with regular spectra. The perception of simultaneously-presented vowels or a mistuned harmonic in an otherwise periodic complex sound have served as models for cognitive studies on concurrent source sound segregation. From previous studies of pitch perception we know that a slightly (about 1-2 %) mistuned harmonic may still contribute to the pitch determination of a complex sound, whereas a component mistuned by 10-16% is usually perceived as a separate, stand-alone sound (Dyson and Alain, 2004). Neural mechanisms for detecting spectral regularities have been conceptualized as harmonic sieves or templates (Duifhuis, 1982, Meddis and O’Mard, 1997) and may involve neurons sensitive to equal spacing between concurrent tones. Two simultaneous vowels are progressively better detected when differences between their fundamental frequencies increase (Chalikia and Bregman, 1989).

Dyson and Alain (2004) measured middle-latency auditory evoked potentials for tuned and mistuned stimuli and found that the Pa wave at 30 ms was significantly larger when a harmonic was mistuned. The increased Pa amplitude was also associated with an increased likelihood that participants would report the presence of multiple concurrent sound objects. Alain and coauthors (2005) recorded ERPs while participants were presented with a mixture of two phonetically different vowels. They analyzed the data with respect to participants' performance. The earliest ERP modulation that distinguished between the trials on which one vowel was perceived as opposed to two concurrent vowels, was a negative wave that was superimposed on the N1 and P2 waves and peaked around 145 ms after sound onset. This component was maximal over midline central electrodes. In similar experiments that utilized mistuned harmonics of a complex sound, a similar pattern of responses has been obtained (Alain, 2001). This led to the proposal that this potential reflects an early stage of auditory scene analysis in which acoustic properties such as mistuning act as preattentive segregation cues that can subsequently lead to the perception of multiple auditory objects. It was proposed that this component should be called 'object related negativity', ORN.

A similar pattern of results was obtained in a study that used spatial location as a cue for sound segregation (McDonald and Alain, 2005): object related negativity was present even when the tonal component was segregated based on a location cue. Johnson and coauthors (2003), as well as Chait and coauthors (2006) found that the perception of dichotic pitch[1] along with the background noise was also paralleled by a negative deflection, similar to object related negativity. Perception of concurrent auditory objects has also been associated with a late positive wave that peaks at around 400 ms following stimulus onset (300-600 ms after sound onset). The amplitude of this component increases with the amount of mistuning of a harmonic, but only if participants are engaged in another task, and the magnitude of the response is affected by contextual manipulations, such as the probability of a mistuned harmonic within a sequence of the stimuli (Alain et al, 2001, Alain et al, 2002). It has been proposed that this 'P400' response reflects conscious evaluation of the sound mixture in which context and memory are used to guide the response (Alain, 2007). This proposal has been confirmed by another study in which dichotic pitch perceived in a noise background also triggered a P400 (Hautus and Johnson, 2007).

In summary, perception of concurrent sounds appears to be dependent on early sensory mechanisms, presumably linked to stimulus-specific representation of the sounds, as well as later stages of processing, the neural correlates of which remain to be uncovered.

---

[1] Dichotic pitch is a pitch percept that occurs when a binaural delay is introduced to a narrow frequency region of the same noise segment presented through headphones binaurally.

CHAPTER 3. Perceptual restoration of missing sounds in noise

Take care of the sense, and the sounds will take care of themselves.

L. Carroll, Alice in Wonderland

The effects of masking are counteracted by the ability of the brain to infer missing sounds from the context of an auditory scene, and thus to perceptually "restore" the integrity of the auditory objects. The greatest advantage of this process is its ability to cope with perceptual uncertainty. Unless the background interference is very strong, we are consciously aware only of the most plausible solution the auditory system has come up with. For example, if one of the phonemes of a continuing sentence is replaced by a noise whose structure favors perceptual restoration (one that plausibly masks the phoneme), listeners report that the word containing that phoneme has been perceived intact, and have problems with reporting which of the phonemes was replaced by noise (Warren 1970, 1984, 1999,2008).

Counterintuitively, addition of noise can also improve speech recognition (Warren, 1999, 2008), a finding which has subsequently been used in robotic auditory scene analysis (Okuno et al, 1998). The known psychophysical phenomena related to perceptual restoration as well as current evidence about their underlying neural correlates are reviewed below.

The auditory continuity illusion

The "continuity illusion" is an illusory perception of sounds as intact, even when parts of them have been replaced by gaps or rendered inaudible by being masked by an extraneous sound.

For example, when silent gaps in a speech sample are filled with high-intensity noise, subjects report the speech as continuing throughout the noise (Miller and Licklider, 1950, Warren, 1970, Samuel, 1981). Although the noise does not add any linguistic information, it improves speech comprehension (Warren, 1999, 2008), due to facilitation of integration of information present in separate segments of the speech (Bregman 1990). Illusory continuity has also been demonstrated for a variety of non-speech sounds, for example pure tones (Warren, 1972), frequency-modulated sweeps (Ciocca and Bregman, 1987, Kluender and Jenison, 1992), and amplitude-modulated sounds (Lyzenga et al, 2005). Apparently, the illusion can be induced whenever a masking sound is loud enough and contains energy at the frequency of the restored sound (Houtgast, 1972, Thurlow, 1957, van Noorden, 1977, Warren, 1984).

This ability to "restore" missing or obscured information is common to many species. Behavior consistent with illusory continuity has been demonstrated in cats (Sugita, 1997) and birds (Braaten and Reynolds, 1999). Cotton-top tamarin monkeys that usually vocalize in response to conspecific calls also respond to a modified version of the call, in which a portion of the sound has been substituted by a noise burst, but not to the noise itself or the incomplete call (Miller et al, 2001). Macaque monkeys judge a discontinuous tone with the gap filled by masking noise as continuous even when such a judgement is punished by a lack of reward (Petkov et al, 2003, 2007)

Perceptual restoration is not unique to the auditory system, because it is found in other sensory systems as well. For instance, in the realm of vision, color, brightness, texture and motion can all be interpolated across a blind spot or a scotoma (reviewed in Komatsu, 2006). The resulting “filling-in” is so strong and realistic that many patients with retinal lesions are not aware of this fact until a medical examiner brings it to their attention. In this case, the “filling-in” appears to be the result of a reorganization of receptive fields of the neurons that, in the absence of input from the lesion region, integrate the information from the surrounding areas. Another example of visual filling-in is perceptual completion of occluded objects. Covert expectations about the properties of filled-in features contribute heavily to object recognition in the visual system (Weerd, 2006; Meng et al., 2005; Ramachandran and Gregory, 1991) and can guide our actions. To maintain perceptual continuity of the visual stimuli during saccades, the brain extends the percept of the saccadic target backwards in time to just before the onset of the saccade. This effect is critically linked to perceptual mechanisms that identify a target's spatial stability (Yarrow et al, 2001).

If there are any differences between the neural mechanisms for filling-in in the visual and auditory domains, these are likely to be due to the nature of the input. Spatial interpolation is readily available for the visual system due to the spatially arranged input on the retinal photoreceptor sheet. The analogue in the auditory system is the tonotopically-arranged input from the cochlea, where the interpolation is likely to be spectrotemporal. Filling-in in both the time and frequency domains may be crucial for maintaining perceptual integrity of auditorily-defined objects. Perceptual restoration mechanisms are very sensitive to binaural cues (for example, a gapped sound does not appear to be continuous if the masking noise is presented dichotically into another ear (Kashino and Warren, 1996)). They also contribute to spatial localization by counteracting the effects of noise. If a tone is presented unilaterally, and the co-occuring noise bilaterally, the tone appears to originate from a central location rather than being localized as lateral to the head (see Warren, 1999 for a review). Predictability of the sound changes in time and frequency can thus be used to reconstruct target signals even though some parts of them are masked by other sounds.

Interestingly, auditory filling-in in the time domain resembles extrapolation more than interpolation: the percept appears to be dominated by the portion of the sound that occurred before, rather than after, the interruption. For instance, if the frequency modulation rate or amplitude modulation rate differ before and after a noise burst, the modulation-rate transition is perceived to occur near the end of the noise burst (Lyzenga et al, 2005).

Psychophysical studies on the continuity illusion led Bregman (1990, p.p. 349-369) to formulate several factors that influence the strength of the continuity illusion:

(1) The ‘no discontinuity’ rule postulates that there should be no evidence that the noise covers silence rather than the continuation of the sound. For example, spectrally dissimilar noises or noises presented in another ear do not cause continuity illusions because the gap in the tone is clearly audible in these cases. A silent gap (typically 50-ms long) inserted between the tone and a masking noise also disrupts the illusion.

(2) The ‘sufficiency of [neural] evidence’ rule suggests that ‘if the peripheral units [of the auditory system] stimulated by a louder sound include those which would be stimulated by the anticipated fainter sounds, then the fainter sound may be heard as present’ (see also Warren, 1972).

(3) The ‘grouping’ rule (Bregman 1990, p.352) states that the pre-noise and post-noise portions of the tone should be grouped into a single auditory object.

(4) The ‘perceptual segregation’ rule (Bregman 1990, p.369) says that tones and noises should be perceived as 2 distinct objects. If they merge into a single stream the transition from one to another rather than continuity will be heard. A smooth transition between the target and the masker favors their grouping in one perceptual stream and thus no restoration occurs. For the illusion to occur the target should be well-established and predictable.

(5) The 'perceptual stability rule'. Changing the tone just before or after the occurrence of the noise disrupts illusory continuity (Bregman and Dannenbring, 1977). Drake and McAdams (1999) examined the strength of the illusory continuity of fainter tones interrupted by louder tones with respect to the relative duration of the fainter and louder tones. Illusory continuity occurred only if the duration of the fainter tones exceeded the duration of the louder tones by at least a factor of 2.

The neural mechanisms of the auditory continuity illusion are currently unknown, even if there is no shortage of hypotheses to explain them. All current hypotheses rely on the idea that, at some stage of processing during the noise, the illusory sounds result from the presence of some kind of continuous brain representation similar to what would be found if the sounds were indeed continuous. Neurophysiological data consistent with such hypotheses are discussed below (Sugita, 1998, Petkov et al, 2007). Neural evidence consistent with the possibility that the sound is continuing behind the noise, and the conservative maintenance of the previous auditory grouping scheme (that is, that the current percept is explainable using an 'old-plus-new' heuristic proposed by Bregman, 1990) appear to be the two most important logical prerequisites for the continuity illusion.

Thurlow (1957) proposed that in case of illusory continuity, the masker stimulates the same neurons that respond to the target sound. Houtgast (1972) suggested that illusory continuity of tones may occur when there is ‘no (perceptible) increase in neural activity’ at masker onset. This point of view implies that continuity may be heard only if a masker has sufficient spectral density to fully mask the target sound.

Experimental studies disagree on the relative spectral densities of the masker and the target that are necessary for the continuity illusion to occur. Warren et al (1972) compared the levels of noise that cause masking and that cause a continuity illusion. They concluded that the masking thresholds for the illusion were smaller than those for illusory continuity. In human psychophysical experiments, noise added to speech at levels of as little as 30 dB below the peak level of the speech enhances intelligibility (Warren et al, 1997). The level of the noise which is most beneficial for restoration of the intelligibility of highly interrupted speech is 10 dB below the peak level of the speech. Thus, at least in humans, full substitution of energy present in the target by energy in the noise is not a necessary condition for the continuity illusion to occur. *Thus, in humans, illusory continuity may result from other mechanisms in addition to the continuous stimulation of peripheral neurons*. In birds, however, only an addition of a noise which is 20 dB louder than species vocal signals and is spectrally confined to a single auditory filter width results in both behavioural responses consistent with perceptual restoration, and sustained activity of auditory systems neurons during occluding noise (Seeba and Klump, 2009, Klump, personal communication), which may point towards interesting differences underlying auditory perceptual mechanisms between birds and primates.[2]

Sugita (1997) recorded activity of the neurons in cat auditory cortex that responded to a frequency glide. If a gap was introduced into the glide, their firing rate decreased during the gap. The addition of a masking noise burst at a sound level that diminished the performance in a gap detection task in cats and human subjects produced a notable *increase* in neuronal firing rate during the noise, even though these neurons did not respond to noise presented in isolation. Similar results have been obtained by Petkov and colleagues (2007) in macaque monkey auditory cortex. During continuity illusion conditions, 74 percent of neurons recorded in the primary auditory cortex demonstrated responses that were similar to responses to a continuous tone rather than to noise presented in isolation. Neurons with phasic responses to tone onsets fired only for the first onset of the tone and did not change their activity with the appearance of the second portion of the tone after the noise was finished. Similarly, offset responders did not fire during the appearance of the noise burst as if there would be no tone offset present. Sustained activity neurons responded throughout the noise even though some of them did not respond to the noise

---

[2] Alternatively, the differences between these groups might be an artifact of the lower target tone loudness used in the avian experiments. The effect of target loudness on the strength of the continuity illusion has not previously been investigated to our knowledge.

presented alone. The authors concluded that the mechanisms responsible for the illusory continuity of pure tones in noise might already be completed at the level of primary auditory cortex. In this study, the short gap in the stimulus tones was masked by a much louder white noise burst (white noise with an upper cutoff frequency of 25 Khz was presented at a level of 63 dB SPL, whereas the tones were presented at 45 dB SPL, 18 dB softer than the noise; gap duration (including ramps to avoid transients) was 56 ms).

When the target tone and a sufficiently-loud masker tone are simultaneously present, the activity of peripheral neurons may ambiguously be attributed to the target as well as to the masker sound; the two sounds could therefore be attributed to separate auditory objects based on the contextual evidence that occurred previously to the onset of the masker tone. This will be true regardless of whether the tone continues behind the noise or not. Warren et al. (1972) suggested that a portion of the internal representation of a masker may be subtracted, or ‘reallocated’ to serve as a substrate for the “active” synthesis of the apparently continuous fainter sound. In the case of the apparent continuity of softer tones caused by louder maskers, the latter appear to be perceived as softer and to have different timbre compared to the same tones in absence of the restoration, consistent with this ‘reallocation’ hypothesis of illusory continuity (Warren et al, 1994, 1999, Drake and McAdams, 1999). In the case of phonemic restoration, where speech signals are restored by a louder masking noise, the perceptual qualities of the masker do not change (Repp, 1992), suggesting perceptual restoration of complex sounds may occur independent from processing perceptual qualities of the masker.

Complex mechanisms of grouping and segregation seem to be involved in perceptual restoration of complex sounds

The ‘gap transfer’ illusion is another example of perceptual restoration, in which the presence of a louder sound that would mask the target is not required. When a long glide with a short temporal gap in its middle crosses with a continuous short glide at the temporal midpoint of both glides, the gap is perceived in the short glide instead of in the long glide (Figure 3.1, Nakajima et al, 2000, Remijn and Nakajima, 2005, Kanafuka et al, 2007). A similar effect, called the ‘shared gap’ illusion does not require any physical energy being present during the gap. The longer glide may be perceived as continuous even if a long (2000 ms) and a short (400 ms) FM sound share a gap (<40 ms) at their crossing point (Remijn et al, 2007). When presented in isolation the gap is clearly perceived in each of the FM sounds. When the gaps in the two sounds coincide, however, the gap is commonly perceived only at the shorter sound.

Nakajima and coauthors (2000) hypothesized that the gap transfer occurs through the binding of sound edges that are closest in spectrotemporal proximity. Edge binding might occur regardless of whether the edges belong to the same auditory object or not. If so, in the gap transfer illusion the gap-delimiting edges of the long sound bind to the shorter sound. These observations point to the importance of the interaction between sensory memory and perceptual interpretation of sound offset responses.

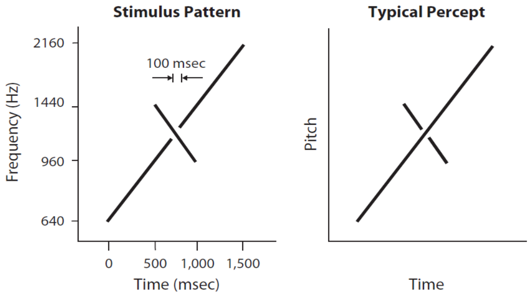


Figure 3.1 The gap transfer illusion (from Kanafuka et al, 2007). When listening to the stimulus pattern (left), the gap is perceived in the short pitch trajectory (right), although it is physically present in the long glide.

Grosberg et al (2004) proposed a common model for both streaming and the continuity illusion which is based on the top-down modulation of excitatory connections from a network layer representing pitch categories. In this model, neurons with on-frequency, off-surround inhibition are balanced with excitatory input such that top-down ‘priming’ inputs sensitize but do not fire target cells, yet matched bottom-up ‘sensory’ inputs and top-down ‘priming’ inputs temporarily amplify the gain of the activities of cells that belong to a perceptual stream. After matching selects consistent components, the continued reciprocal action of bottom-up and top-down inputs generates a "resonance" that is hypothesized to give rise to an auditory percept. The spectral components inconsistent with the stream are freed to be captured by other streams (as in the ‘old plus new’ heuristic of Bregman, 1990). Thus the network can be involved in multiple "resonances", each selectively grouping together those pitches whose frequencies correspond to distinct auditory sources. Behavior of this model is consistent with several aspects of auditory continuity and streaming, including the Deutsch scale illusion (Deutsch 1974, 1975). The assumption of this model is that modulatory feedback is used by the brain to generate a link between a low-level (spectral) and a higher-level (pitch) representation of sounds, and distances at this higher-level representation are then used for streaming purposes. The existence of strong feedback projections from the auditory cortex to virtually all subcortical structures (Winer, 2006), and evidence for short-term modulation of the thalamic activity by auditory cortex (Yan and Suga, 1998), makes it likely that feedback projections play some kind of role in auditory processing and the formation of auditory objects. However, there is no direct evidence supporting the particular role hypothesized by this model.

Another model by Husain and colleagues (Husain et al, 2005) utilized a large-scale model with multiple areas connected by feedforward and feedback projections. The authors systematically varied the strength of the connections and concluded that the feedforward connections between primary and secondary auditory regions of the model were necessary for the continuity illusion, which “is a network property mediated by the increase in the spectro-temporal receptive fields from the primary to secondary auditory cortex to tertiary processing areas in the superior temporal cortex“. Though results of several studies are consistent with increasing spectro-temporal receptive fields (STRFs) along the cortical hierarchy (see Hussain and Horwitz (2006) for a review), currently there are no electrophysiological studies that have confirmed the basic tenets of this model for the continuity illusion.

Illusory continuity as a form of sensory memory

Several aspects of the continuity illusion point toward its tight relationship with auditory short-term memory (‘sensory’, or ‘iconic’ memory).

(1) First, during the continuity illusion (sounds that appear to continue in the presence of a noise), the listener hears exactly the same sound as before the noise (same pitch, amplitude and frequency modulation- Lyzenga et al, 2005) – a property resembling the so-called ultra-specificity of iconic memory (Gabrieli, 1998).

(2) Second, the time constants of the continuity illusion are similar to those reported for iconic memory (from tens of milliseconds to seconds). In the original reports, the illusion spanned 50-100 ms (Miller & Licklider, 1950; Warren, 1970), which was however extended up to 10-15 seconds for noisy targets (Warren et al, 1972) and up to several seconds for tonal targets (Riecke et al, 2008, and the present experiments, see below). These time constants are within the reported span of auditory short-term memory (Samms et al, 1993).

(3) Perceptual restoration of some sound features seems to be constrained by the properties of the underlying neural mechanisms. For example, Crum and Hafter (2008) investigated perceptual completion of an upward frequency glide through the noise. They found that the velocity of the frequency change of the illusory part of the glide is smaller than that of the initial part of the glide. The participants were asked to adjust the starting point of the second portion of the glide so that the perceptual integrity of the glide would be maximally preserved. They consistently adjusted the second portion of the tone so it would start from a lower frequency than expected if the extrapolation would be linear. Such results might be due to changes in an object’s feature representation under conditions of perceptual uncertainty. If the motion direction and velocity are represented by a population of neurons with individual members that are tuned to different velocities, then one active neuronal subpopulation could inhibit the other, currently inactive subpopulation by a winner-take all mechanism (Razak and Fuzessery, 2008). The contrast in neuronal firing rates between the active subpopulation and the rest of the network might gradually decrease due to a lack of specific input during the masking or occlusion period. The resulting decrease in the strength of the velocity representation over the whole network would result in a decrease in the perceived velocity of the frequency change. An alternative explanation for the results of Crum and coauthors could be that the decrease in the frequency modulation rate

results from the fact that the second portion of the sound was not a glide but a steady pure tone, so the auditory system could have maximized trajectory smoothness to account for a sudden change of the trajectory of the tone.

Sensory memory for individual object features has generally been proposed to be stored in the same parts of brain networks that represent those features. For example, the secondary visual area MT, which contains neurons that are highly sensitive to object motion, is active during perceptual filling-in of apparent motion (Liu et al, 2004). Primary visual cortex V1 is also involved in the representation of apparent motion, and BOLD responses in V1 to apparent motion cues are present in subareas that are not directly activated by apparent-motion-inducing stimuli, suggesting the possibility of a feedback interaction between MT and V1 in apparent motion representation (Muckli et al, 2005, Larsen et al, 2006). Similarly, when pure tones are used to elicit an auditory continuity illusion, a BOLD response is seen in the middle part of Heschl's gyrus (Riecke et al, 2007). Changes of neural responses to acoustically identical stimuli parallel changes of listeners' reports of perceived continuity for these same stimuli, thus confirming that these responses are linked to subjective percepts. As in the case of vision, feedback interactions with higher-level regions may be recruited during filling-in of more complex sounds such as vowels: activation in vowel-sensitive regions of the middle temporal gyrus was greater in the illusion condition than when the illusion is prevented by the introduction of a small silent period between the vowel and the noise (Heinrich et al, 2008). Activity in Heschl's gyri (HG), the approximate location of the primary auditory cortex, showed the opposite pattern, so the authors concluded that it may depend instead on the number of perceptual onsets in a sound. However, it is also possible that the changes in HG activation are reflecting changes in feedback relations among auditory regions, rather than changes that are directly tied to context-independent physical features of the sounds.

Shahin and coauthors (2009) performed an fMRI study in which subjects listened to words with brief noise bursts of varying duration centered about fricatives or affricates. The white noise segment either replaced or was superimposed on the speech sound. The authors suggested that the continuity illusion is based on two neurally dissociable processes: a repair process which restores the low-level sensory representation if primary sensory information is degraded or missing, and a “subjective continuity process” where the structural coherency of the stimulus is evaluated. They classified regions that showed greater activity when less sensory information was available as “repair-mediating” (bilateral anterior insula, left inferior frontal gyrus, pre-supplementary motor area). Areas that were more active during the continuity illusion than when the illusion failed were classified as “subjective continuity” areas (left angular gyrus and superior temporal sulcus (AG/STS), right posterior superior temporal sulcus, bilateral superior frontal sulcus, and precuneus). The core of their continuity network, left AG/STS, displayed significant specificity for words versus pseudowords. Interestingly, some of the areas in which activity changed along with perceptual judgements in this study have been hypothesized to belong to a 'default-mode network', in the brain which reduces its activity upon commencement of goal-directed, attentiondemanding tasks (Greicius et al., 2003 and Raichle et al., 2001). In these areas, BOLD responses were less suppressed during the illusion than during illusion-failure (Shahin et al, 2009). Consistent with earlier studies (Lyzenga et al, 2005) the authors suggest that perceived continuity could be "... partly mediated by passive processes, for glossing over interruptions, as much as active ones".

In the absence of studies that have directly assessed sensory memory contributions to the continuity illusion, it can be suggested that examining “sensory memory” contributions may provide a fruitful avenue for future studies on both “passive” and “active” mechanisms of the continuity illusion.

CHAPTER 4. A conceptual neural mechanism underlying continuity-illusion type effects

It's a poor sort of memory that only works backward.

L. Carroll, Through the Looking Glass

In the search for continuity illusion mechanisms, it is important to consider two main potential prerequisites for the formation of auditory objects. One is stimulus-specific, short-term changes in the responsiveness of neural circuits; the second one is “grouping by neural synchrony”, in which sounds that produce temporally uncorrelated neuronal responses are hypothesized to belong to different auditory objects. In this framework, a numerical thought experiment was used to aid deliberations about possible mechanisms underlying the auditory continuity illusion for pure tones and band-limited noise maskers. The aim of this exercise was to consider the simplest possible mechanism that might account for the basic features of the continuity illusion.

The neural representation of an ongoing tone undergoes time-dependent changes (which also may yield a 'sensory memory' for a time interval, see Buonomano and Maass (2009) for a review). These changes potentially allow someone who is recording from those neurons to differentiate between ‘previously stimulated’ and ‘not-previously stimulated’ neurons, when a noise starts sometime after the ongoing tone. Population increases in phase coherence of auditory cortical neurons have been proposed to underlie the perceptual formation of new auditory objects (reviewed in Elhilali et al, 2009). The onset of a noise after a tone causes a population onset response in the auditory cortex (Heil, 2001), which may be the result of temporal synchronization of the firing pattern of the neurons that have previously been silent (“grouping by shared input”). One could hypothesize that those neurons whose best frequency matches that of the target pure tone stimulus form a phase-coherent “response group” that is different from the “response group” formed by the previously-silent neurons that respond in a phase-coherent fashion when the noise masker starts. The formation of such dynamically segregated response groups in primary and/or secondary auditory cortices may be a neural prerequisite for the perceptual occurrence of auditory streaming and the perceptual continuity of the tonal target stimulus.

Spike rate adaptation is one of the most commonly found neuronal correlates of sensory memory formation in the auditory system. The circuit and cellular mechanisms (including the intrinsic membrane and channel properties of constituent neural cells) that underlie habituation of neuronal responses may be quite varied among different groups of neurons, and it is not entirely clear which of these are operating in the auditory system. Whatever mechanisms are actually responsible for the origin of time-dependent segregation of neuronal responses, their behavior can be conceptually captured by a simple network model that exhibits consistent changes in neuronal response during the presentation of the sounds. An example of such a network (which uses short-term synaptic facilitation) is presented below. Similar examples can be built using synaptic depression (which would result in spike rate adaptation with time, much more commonly seen in neural recording than facilitation of responses). However, there is good evidence for the involvement of synaptic facilitation in early sensory processing, which comes from the observation of paired-tone facilitation in the cochlear nucleus (Palombi et al, 1994).

A single mechanistic process with a time scale that is long enough to segregate the responses to the tone and the noise is required; however, it is also necessary for this process to tolerate small gaps between the tone and the noise which are known to disrupt illusory continuity (reviewed in Warren, 1999). The latter condition could be met if the "memory" for the immediate previous history of stimulation is somehow "reset" during the gap. This is accomplished below by including two equations that define synaptic weight dynamics: one of them (4.3) describes the increase in synaptic weights with an increase in firing rate. The second component (4.4) is dependent only on the synaptic weight itself, and assures that it can decay rapidly in the absence of the stimulation. The incorporation of this second time constant is in line with previous suggestions by Bregman and colleagues that both illusory continuity and streaming depend on a common preliminary mechanism that links together the parts of a sequence that have similar frequencies, but that the time constants for these two processes are somewhat different (Bregman et al, 1999).

These constants were adjusted to illustrate the idea that a model based solely on these mechanisms is worth considering. The incorporation of empirical values from electrophysiological recordings (and a much more rigorous formulation of the model) would be required for any serious further development of this idea beyond the thought experiment stage.

The network model used here consisted of two tonotopically-arranged layers of neurons, an input layer and an output layer (Figure 4.1). Neurons in the input layer received frequency-tuned input. Only direct feed-forward connections were present between the input and the output layers.

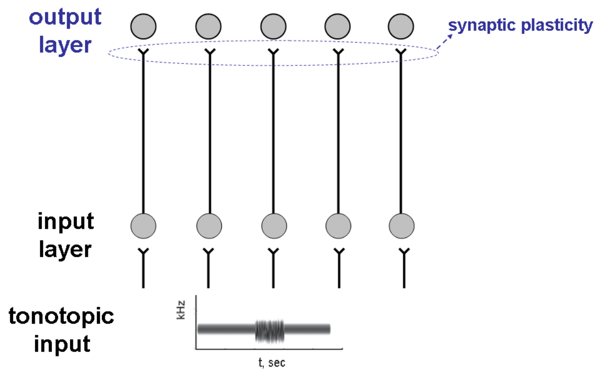


Figure 4.1 Architecture of the model.

Short-term synaptic plasticity was implemented as an increase in the connection strength between the corresponding neurons in the layers, instantiated in a matrix of synaptic weights. In the absence of stimulation, synaptic weights rapidly regressed to a default value. The firing rate of neurons in the model depended on their previous firing rate and the amount of input received during a given time period:

$$\tau_m \times \frac{du}{dt} = -(\ u - u_{res}\ ) + r_m \times I_{in} \times w \qquad (4.1)$$

where *u* = firing rate (spikes/sec), u_{res} = resting firing rate (spikes/sec), w is a synaptic weight, I_{in} is the synaptic input (μA), τ_m is the membrane time constant (20 msec), r_m is the membrane resistance (1 MΩ), and *dt* is 1 ms.

Synaptic weights depended on the firing rate of the presynaptic neuron only (short-term synaptic facilitation, which can be mediated by, for example, residual calcium level increases in the synapse (Zuker and Regerh, 2002)), and also decayed with time:

$$w_i = w_{i-1} + dw_{(w)} + dw_{(u)} \qquad (4.2)$$

$$\tau_{synaptic\_facilitation} \times \frac{dw}{dt} = u \times \left(w_{\lim\_upper} - w_{i-1}\right) \qquad (4.3)$$

$$\tau_{synaptic\_decay} \times \frac{dw}{dt} = w_{\lim\_lower} - w_{i-1} \qquad (4.5)$$

where τ indicates the time constant ($\tau_{synaptic\_facilitation}$=600, $\tau_{synaptic\_depression}$=10), and *w* indicates the synaptic weight. $w_{\lim\_upper}$ and $w_{\lim\_upper}$ are the limits for the synaptic weight value (600 and .2 respectively).

The neurons whose best frequency matches that of a tone responded differently to the noise when compared to the neurons that were not stimulated by the tone (figure 4.3 A; figure 4.3 C). If a gap is introduced before the noise, the responses become more similar to each other due to the decay of the dynamically-induced tone-stimulus-specific response changes (figure 4.3 B, figure 4.3 D)

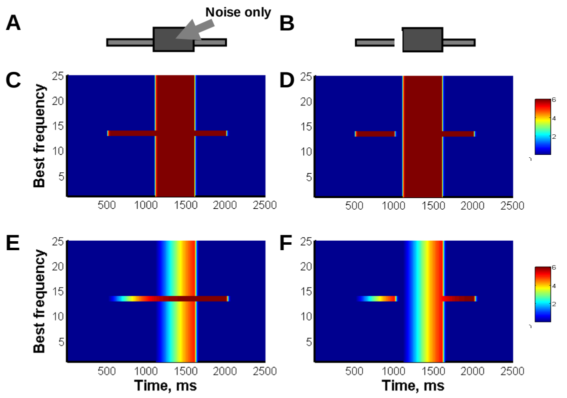


Figure 4.2 Stimuli, input and output layer activity. A, B – Schematic representation of the stimuli in a continuity illusion condition (A) and in an acoustically similar condition that prevented the occurrence of the continuity illusion (B). (C ,D ,E ,F) - Firing rate of the neurons (spikes/sec) as a function of their best frequency and of time. (C), (D) – Activity in the input layer. Note that in both conditions, the firing rate of the neurons responsive to the tone is indistinguishable from other neurons during the noise. (E), (F) – activity in the output layer. In the illusion condition (E), the firing rate of the neurons responsive to the tone is higher than that of the other neurons during the noise. Moreover, it is similar to the time course of the response to a pure tone without the noise (not shown).

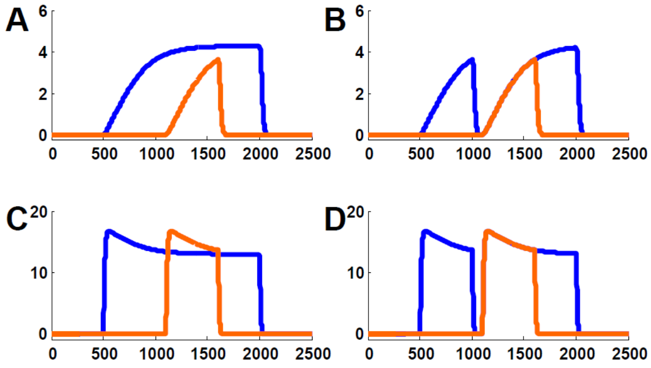


Figure 4.3 Simulated peristimulus time histograms of a neuron with a best frequency that matches a tone (blue) and is off the frequency of the tone (orange). A,B – the “synaptic facilitation” version outlined above (eq. 4.1-4.5 and figure 4.2). C,D – a neuron with spike rate adaptation (data obtained by using the first neuron as inhibitory). A,C – responses to a continuity illusion stimulus (figure 4.2 A), B,D – responses to a stimulus with a gap (figure 4.2 B). Note that the responses of the two neurons are identical in the case of a gap but are dissimilar during the noise in the case of the continuity illusion.

This simple thought experiment adds something to the original explanation of Thurlow and Elfner (1959), which stated that a masker stimulates the same neurons that respond to a target sound along the temporal dimension. The masker does stimulate the same neurons, but the dynamic changes in their responsiveness induced by the preceding tone stimulation allows them to represent information about the presumably-continuing duration of the target sound, and enables the continued segregation of this response group of neurons from new responses groups that could be formed as new components get added to the sound scene. This is presumably the neural instantiation of Bregman’s (1990) ‘old plus new’ heuristic.

This thought experiment stresses the importance of two different time constants for the continuity illusion: the time constant of adaptive responsiveness changes in the neurons that respond to the tone and those that respond to the noise, and a separate time constant expressing how fast these changes in responsiveness decay back to their initial values when the stimulation is no longer present. It makes intuitive sense that the decay of responsiveness changes should be faster than the induction of the response changes during the tone, because a relatively fast decay would allow the auditory system to detect short gaps that may signify the edges of sounds, and would explain why the auditory continuity illusion will not span across gaps. The relatively more sluggish rate of change of those neurons that respond to the tone and the noise should, in principle, determine the maximum duration of the illusion. If the latter adaptation rate is fast, the firing rate of the neurons responsive to the noise will soon match the firing rate of the neurons that have been responding to the tone. If the adaptation rate is slow, the firing rates of the neurons responsive to the tone will be different from those responding to the noise for a longer period of time.

This simplistic thought experiment is also consistent with time-dependent changes in masker perceptual qualities (loudness and timbre) in the case of the continuity illusion (Warren et al, 1999, Drake and McAdams, 1999). Such behavior may also be related to the recently described “duration illusion” (Carlyon et al, 2009), in which the perceived duration of a narrow-band noise can be increased by immediately preceding wide-band noise of a similar frequency. The maximal increase in perceived duration is about 60–80 ms. The effect typically disappears when a 20–40 ms gap is introduced between the wide-band noise and the narrow-band noise, even if for some participants the illusion can persist at longer gap durations (80 ms, Carlyon et al, 2009). In principle, such lengthening of the perceived duration of the narrow-band noise can be explained by the adaptation of the neurons to the wide-band noise; these lower firing rates would be consistent with a sound that has been going on for a longer duration during the narrow-band noise stimulus.

In summary, many different processes may lead to the segregation of neural responses to previously existing and novel sounds. The thought experiment sketched out above examined the role that duration-dependent neural processes might play in the segregation of such responses. The thought experiment suggests that it is clearly advantageous to incorporate neural phenomena such as spike rate adaptation and short term synaptic plasticity (adaptation of facilitation) into future thinking about the underlying mechanisms for the segregation of groups of neurons in auditory scene segregation. This may also provide a mechanistic basis for “grouping by coactivation” and the assignment of distinct patterns of activity to distinct auditory objects (illusory continuous tone and noise). Mechanisms of duration-dependent changes in the neural representation of sounds may therefore constitute one of the basic mechanisms involved in the neural instantiation of the continuity illusion for simple steady sounds.

CHAPTER 5. Illusory discontinuity

The Mock Turtle: Well, I never heard it before, but it sounds uncommon nonsense.

L. Carroll, Alice in Wonderland

The experiments reported in this chapter stem from an unexpected observation made during a pilot experiment for studying EEG correlates of the auditory continuity illusion. The briefing procedure for this experiment included examples of continuous and discontinuous tones overlaid with noise bursts, which were presented to participants prior to in the start of a psychophysical experiment. Surprisingly, a few subjects spontaneously reported that the example of the "continuous" tone (accompanied by a non-masking noise burst) did not sound entirely continuous to them. In fact, during the subsequent psychophysical test, they tended to mistake continuous tones for discontinuous ones in the presence of short noise bursts that were far from the frequency of the continuous tone. They described their perceptions as either a subjective gap in the tone, or as a difficulty following the tone during the noise burst.

The two experiments described in this chapter were undertaken to better characterize variation among subjects in their susceptibility to perceiving "illusory discontinuity", to document the relationship this bears to individual susceptibilities to perceiving the auditory continuity illusion, and to document the relationship between illusory discontinuity perception and auditory streaming based on non-spatial frequency separation cues.

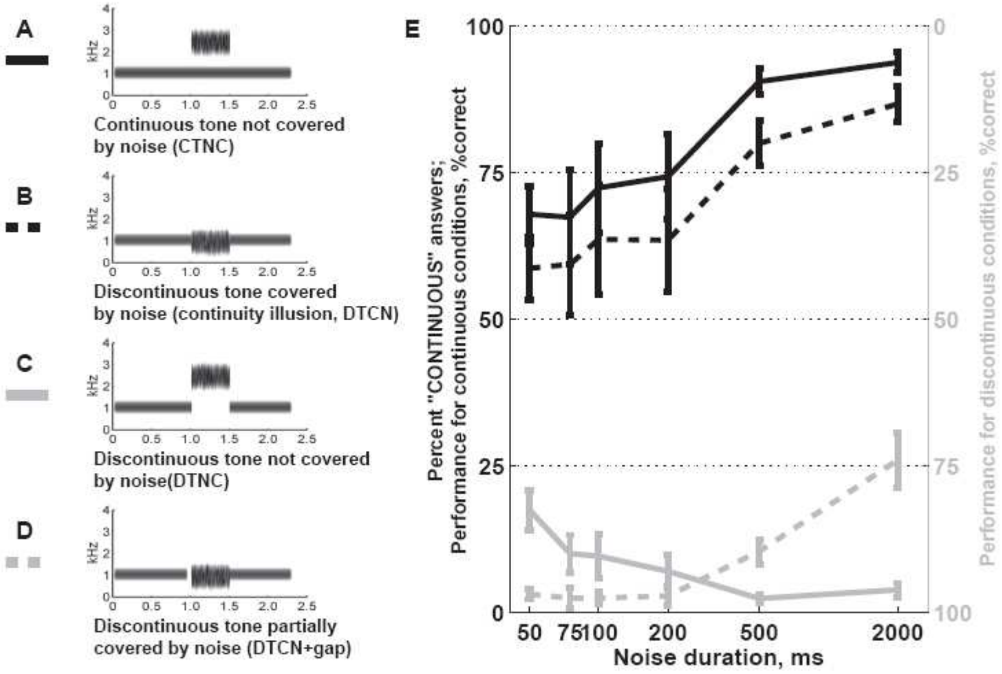
Methods: Experiment 1. Perception of tones in noise

A. Participants

Fifty-four subjects were recruited from the local population in Trieste, Italy. All subjects were naive to the purposes of the experiment and were paid an hourly wage. Their ages ranged from 18-32 years (mean ± sd, 23.3 ± 3.5 years). The subjects reported neither personal histories nor a family history of neurological or psychiatric disorders, nor any hearing problems. Data from 8 subjects were excluded from further analysis (1 because of a large interaural hearing level difference, 1 because of poor performance in catch trials in the psychophysical tasks, and 6 who were unavailable to complete all of the data collection procedures); the final subject pool thus consisted of 46 subjects (20 males). Informed consent was obtained from each participant after they were instructed about the experimental procedures. The use of human subjects for these experiments was approved by the SISSA experimental protocol committee.

B. Apparatus

Stimuli were generated using the SIGNAL Digital Programming Language (Engineering Design, Berkeley, CA) (for replication 1), or MATLAB (The MathWorks, Natick, MA) (for replication 2), and played using an M-Audio Transit soundcard, through E-A-RTONE 3A insert earphones (Aearo Corporation, Indianapolis, IN). All stimuli were generated using a 44.1 kHz sampling rate. The experiments took place while participants were seated in front of a computer in a sound-attenuated room. Perceptual judgments were recorded using custom programs written in SIGNAL or MATLAB, and stored for later analysis.



A: Continuous tone not covered by noise (CTNC)

B: Discontinuous tone covered by noise (continuity illusion, DTCN)

C: Discontinuous tone not covered by noise(DTNC)

D: Discontinuous tone partially covered by noise (DTCN+gap)

Figure 5.1 Stimuli and continuity task. **A-D:** Spectrograms of the stimuli. In conditions B, C, and D there is no tone physically present under the noise. However, the tone in condition B may be perceived as continuous (the continuity illusion). In condition D the gap starts 50 ms earlier than in C and B. The lower edge of the noise in conditions A and C are separated from the tone by more than one octave. **E:** Psychometric functions. The Y-axis on the left represents the percentage of ‘Continuous’ answers for conditions A and B. The Y-axis on the right represents the percentage of “Discontinuous” answers for conditions C and D. The mean value ± 1 SE is shown (N=46 for noise durations of 50, 500, and 2000 ms, N=18 for 75,100 and 200 ms). Combined data for pure tones and AM tones are reported here.

C. Stimuli

Figure 5.1 A-D illustrates the four categories of stimuli employed in the experiments. The stimuli were designed to render equal probability of perceiving the tone as continuous and discontinuous: (25% physically continuous trials (Figure 5.1A), 25% illusory continuous trials (Figure 5.1B), 25% discontinuous with non-masking noise (Figure 5.1C), 25% discontinuous with masking noise (Figure 5.1D)). The tone could be either a 1007 Hz pure tone, or the same tone modulated by a 40 Hz sinusoid with 90% modulation depth. On each trial the noise burst (7 dB louder than the tone) started 1 second after tone onset. Two noise types were used: one that could potentially mask the tone (and induce the auditory continuity illusion) with a bandwidth of 600 Hz centered at 900 Hz (Figure 5.1B, D), and a second noise with the same bandwidth centered at 2425 Hz (from 2125 to 2725) (Figure 5.1A, C), which does not mask the tone (since its lower edge is separated from the tone by more than one octave - about 6 equivalent rectangular bandwidths (ERBs) away[3]).

Tones and noises were combined in 4 experimental conditions that can be grouped according to the type of noise they contain. The conditions in Figure 5.1A and C represent well-matched cases of physical continuity (Figure 5.1A) and discontinuity (Figure 5.1C) of the tone, while the conditions in Figures 5.1B and D represent a case of “illusory” continuity (Figure 5.1B) and the most closely-matched discontinuous stimulus that can be created (Figure 1D): a tone and noise identical to that in Figure 5.1B, but with a short gap inserted in the tone 50 msec before the onset of the noise. These four conditions will be referred to by acronyms: “Continuous tone not covered by noise” (CTNC – Figure 5.1 A), “discontinuous tone covered by noise (DTCN -Figure 5.1B), “discontinuous tone not covered by noise with gap” (DTNC - Figure 5.1C), and “discontinuous tone partially covered by noise with gap” (DTCN+gap – Figure 5.1D).

For the stimuli in Figures 5.1B and C, the tone and the noise had temporally complimentary envelopes, with the tone fading out simultaneously with the noise fading in. All stimuli used 7-ms cosine-square ramps at the sound edges. Independently-prepared noise samples were used for each trial. Both types of band-filtered noises had smoothed rectangular spectral envelopes (30-Hz ramps on each side). The noise level in the conditions in Figure 5.1 B and D was set at 68 dB above the hearing level of each subject; none of the participants found this level to be unpleasant. The noise in the conditions in Figure 5.1A and C were set 0.5 dB softer (67.5 dB HL). In a pilot experiment, these levels of noise were perceived to have equal loudness (data not shown). Two different sets of noise durations were used in separate replications of the experiment: 50, 75, 100,

---

[3] The equivalent rectangular bandwidth (ERB) is a measure which attempts to quantify the effective width of human auditory filters. The number of auditory filters separating two frequencies can be calculated by converting the frequencies using the relation ERBnumber = 21.3*log([0.00437* f]+1), where f is frequency in Hz (Hartmann 1998, p. 251).

200, 500, 2000 ms in the first version of the experiment (in which 18 participants were tested) and 50, 500 or 2000 ms in the replication (a new group of 28 participants).

D. Procedure

Participants were instructed to “ignore the noise, attend to the tone and listen for gaps or any kind of changes in the tone”. If they heard a gap or a change in loudness of the tone they were asked to judge the tone as 'Discontinuous'. If the tone did not change, they were to respond that the tone was 'Continuous'. Examples and a practice session were provided to facilitate attention to the tone during the noise. For the first 10 minutes of the familiarization procedure, participants were asked to detect a change in loudness in an amplitude-modulated tone. The change could occur 1000 ms after the tone onset and lasted for 500 ms. In the second part of the familiarization procedure, also 10 minutes long, the task was the same, but a soft burst of noise similar in frequency to the tone appeared at the same time as the possible change in tone loudness. No feedback was provided. The main experiment was divided into three 15-minute-long blocks with pauses between each block. A response window appeared on the computer monitor screen after sound offset to avoid any cross-modal interactions; subjects indicated their response choice, the data were saved, and a new trial began.

Hearing thresholds were obtained prior to the experiment using custom-written programs that employed two types of the stimulus sounds from Experiment 1 (500-ms-long band-pass noises either from 600 to 1200 Hz or from 2100 to 2700 Hz). The participants’ task was to detect the presence or absence of the sound during a given time interval. An adaptive staircase started from a clearly audible sound level and adjusted the amplitude of the noise in a 2 up, 1 down procedure, with step size decreasing over time (step up/step down ratio = .9). The procedure was stopped after 14 reversals, and the noise detection threshold was calculated as the mean value over the last 10 reversals. These thresholds were determined successively for each ear and noise type in random order.

Music experience was assessed by means of a self-administered questionnaire, and scored into three categories according to the number of years of musical education (less than 2 years, from 2 to 7 years, more than 7 years).

E. Analysis

Performance was characterized separately for physically and illusorily continuous stimuli and their matching discontinuous stimuli using a non-parametric criterion-free measure of sensitivity A’, and B''_d, a measure of bias (Grier, 1971). A ‘Hit’ occurred when stimulus 1A or 1B was judged as continuous. A ‘false alarm’ occurred when stimulus 1C or 1D was judged as continuous. The A’ and B''_d measures were calculated as follows:

If hit > fa

$$A' = 1/2 + \frac{(hit - fa) \times (1 + hit - fa)}{4 \times hit \times (1 - fa)}$$

if fa>hit

$$A' = 1/2 - \frac{(fa - hit) \times (1 + fa - hit)}{4 \times hit \times (1 - hit)}$$

$$B''_d = \frac{(1 - hit) \times (1 + fa) - hit \times fa}{(1 - hit) \times (1 + fa) + hit \times fa}$$

To characterize intersubject variability in performance, data from individual subjects were examined. The consistency of responses was estimated by calculating the probability of obtaining a given distribution of answers by chance using the cumulative binomial distribution function. If the probability of obtaining a given distribution by chance was smaller than 0.05 or greater than 0.95, then a subject's answers were considered consistent. Otherwise, the subject's performance was considered to not differ from chance.

The responses to pure tones and amplitude-modulated tones were not significantly different from each other, and were therefore combined for all data presentations in this study. Since response data were obtained in the form of percentages, we used nonparametric statistical tests. Paired-sample sign tests or Wilcoxon matched-pair signed rank tests were used to compare dependent samples, while Wilcoxon rank-sum tests were used to compare independent samples. The Scheirer-Ray-Hare Two-way ANOVA (Sokal & Rohlf 1995) was used to assess cases with two variables that had multiple states. For single variables with multiple states, Kruskal-Wallis ANOVA was used for independent samples and Friedman ANOVA for matched samples.

Methods: Experiment 2. Relationship between illusory discontinuity and streaming

A. Participants

Twenty-eight participants from the 46 participants described in Experiment 1 above performed in this experiment.

B. Apparatus

Same as in experiment 1.

C. Stimuli

To assess individual streaming abilities, a set of sounds were used that could be spontaneously heard either as : (1) a single melody encompassing high and low frequency tones, or (2) 2 distinct melodies, one composed of high tones and the other of low tones (see Bregman (1990) for a review). On each trial a sequence of tone triplets was presented. Triplets had an ABA frequency structure. Previous work with this paradigm has indicated that the frequency difference between tones A and B determines the probability of perceptual segregation on each trial (Bregman, 1990); in general, the greater the perceptual separation between A and B, the easier it is to segregate melodies into two streams. All the subjects could perceive the tones as strongly grouped or as well-streamed on some trials, as frequency separation changed from 52 to 511 Hz. To obtain the closest match to the tonal stimuli used in Experiment 1 above, the frequencies of tones A and B were centered on 1 kHz and were presented at the same loudness level. Each tone lasted 150 ms; inter-tone intervals were 150 ms. Each trial contained one 4-second-long sequence that was faded in and out with a 1-second-long cosine-square ramp. The tones were presented at 61 dB HL.

D. Procedure

On each trial participants were asked to estimate the degree of perceptual separation of lower and higher tones in the melody on a subjective scale from 1 to 5. 'One' meant that a single melody of alternating tones was heard throughout the trial, and 'five' corresponded to a strong perceptual separation into one lower and one higher pattern starting from the beginning of the sequence. 'Three' was used to indicate trials on which the low and high patterns separated during the course of a trial. Examples of highly grouped and highly separated sound sequences as well as familiarization trials were provided. Due to subject availability this experiment was either performed in the same session as experiment 1, or for 7 subjects, 2-4 days after.

E. Analysis

Individual participants were expected to differ in the degree of frequency separation (Δf) above which they reliably heard separated streams (Bregman, 1990). To compare individual measures of susceptibility to noise with this measure of streaming, it was desirable to choose streaming conditions with sufficient intersubject variability. The tonal patterns were perceived as highly segregated by most of the subjects at Δf values of 8 semitones (2/3 of an octave). At Δf values of <3 semitones, most subjects perceived a single pattern. Intermediate Δf values had intersubject variability that was sufficiently large for a correlational analysis. To obtain a single measure of streaming for each subject, data from intermediate frequency separations (3-7 semitones) were averaged. As a measure of illusory auditory discontinuity, performance in the continuous tone not covered by noise (CTNC) condition (Figure 5.1A) with a 50 ms noise duration was used. The Spearman’s rank correlation coefficient was used to assess the relationship between the two variables.

Results

Illusory discontinuity

To assess individual susceptibility to noise participants were asked to judge the continuity of a 1 kHz tone in several conditions (figure 5.1), which included a physically continuous tone presented together with spectrally remote noise, and a discontinuous tone with spectrally similar noise which usually elicits the auditory continuity illusion. Two matched discontinuous conditions were used in catch trials. Behavioral data were analyzed using methods derived from signal detection theory (Grier, 1971).

The distribution of response bias in trials with 50 ms noise was clearly bimodal: only a small number of participants showed no bias (Figure 5.2). People who displayed a bias to continuity exhibited good performance in the continuous tone conditions but made more mistakes in the physically discontinuous tone conditions: if a short gap in the tone coincided with the appearance of a spectrally remote noise burst, the tone was reported as continuous. On the contrary, people who exhibited a bias to discontinuity had high percentage correct scores in discontinuous conditions but poor performance in continuous conditions: these participants tended to mistake a continuous tone for a discontinuous one. Participants with discontinuity bias also perceived the continuity illusion less strongly than subjects with a bias to continuity.

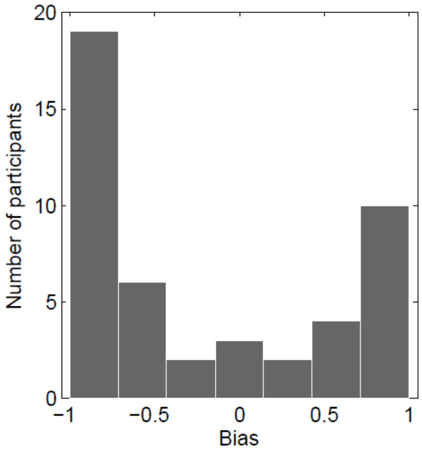


Figure 5.2. Distribution of the individual response biases. '-1' signifies a bias towards 'discontinuous' responses; '1' is a bias towards 'continuous' responses. The data comes from the trials on which a physically continuous tone or a discontinuous tone were presented simultaneously with a 50 ms noise burst off the frequency of the tone. N=46.

There was a significant negative correlation between individual performance in continuous tone and discontinuous tone conditions: the participants who were better at detecting gaps were also more likely to judge a continuous tone as discontinuous (Spearman's $\rho=-0.4$, $p<0.001$, N = 46, Figure 5.4). In some participants (11 of the 46, 24%) the discontinuity bias was so strong that they consistently judged a continuous tone as discontinuous, i.e. they performed significantly worse than expected by chance (p<.05). This phenomenon will be referred to as 'illusory discontinuity of a tone'. The participants that showed the greatest effect described their sensations in terms of the tone actually containing a physical gap. It is possible that in participants with milder disruption of continuity, this phenomenon reflects perceptual uncertainty rather than the actual sensation of a tone being stopped during the noise.

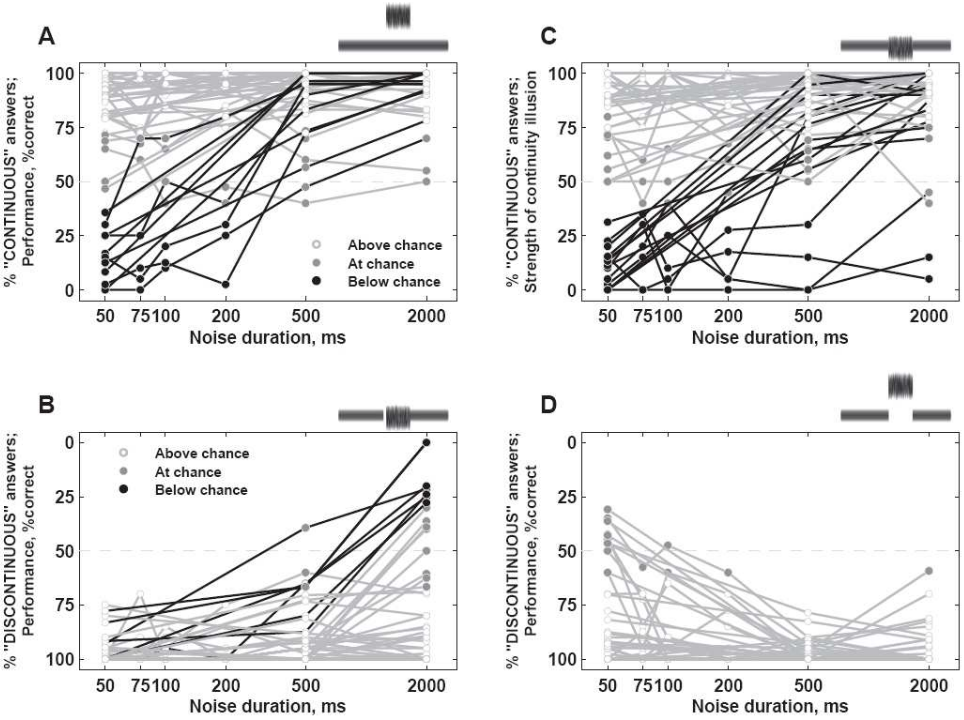


Figure 5.3 Individual performance: Each line represents percent of correct answers given by a single subject in a continuity task at different noise durations. Individual chance levels are slightly different because of a different number of trials. **A. Continuous tone with spectrally remote noise condition** Subjects that consistently judge a continuous tone as discontinuous in trials with 50 ms noise are traced in black and plotted on the foreground. **B. Discontinuous tone with spectrally similar noise condition**. Subjects that consistently judge a discontinuous tone as continuous in trials with 2000 ms noise are traced in black. **C. Discontinuous tone with spectrally similar noise (the continuity illusion)**. Subjects that do not perceive the continuity illusion in trials with 50 ms noise are traced in black. **D. Discontinuous tone with spectrally remote noise** Condition. For all panels, data from 46 participants was used for noise durations of 50, 500, and 2000 ms; data from 18 participants was used for noise durations of 75, 100 and 200 ms.

Group performance in the physically continuous tone with spectrally remote noise condition consistently improved as the noise duration increased (effect of noise duration is significant, χ^2=25.6, N=46, p<0.0001, Friedman ANOVA), i.e. the illusory discontinuity of the tone was not perceived with noises longer that 200 ms. Similarly, gap detection in noise improved with increasing gap duration (χ^2 = 22.3, N = 46, p<0.0001, Friedman ANOVA; group means plotted in Figure 5.1E as a solid gray line, individual performance shown in Figure 5.3D).

The effect of noise duration on the perception of the tone may represent an effect of the length of the integration window available for the perceptual evaluation. It has long been known that a decrease in integration window length leads to diminished spectral resolution and impaired auditory performance in tasks such as pitch discrimination (reviewed in Plack and Oxenham, 2005). The results of the present experiment show that the length of the integration window affects not only bias-free sensitivity, but also influences response bias in the gap detection in noise task. It is suggested that this response bias reflects individual differences in sensory processes that are further assessed in Chapter 6 below.

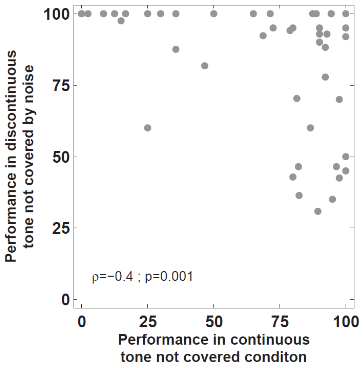


Figure 5.4. Relation between discontinuity illusion and gap detection: performance in the continuous tone not covered by noise (CTNC, Figure 5.3A), plotted on the Y- axis, and performance with a discontinuous tone not covered by noise (DTNC, Figure 5.3D, X–axis); both using data from 50 ms noise durations (N=46).

The measure of bias takes into account the percentage of correct answers in the continuous tone and gap conditions. Filling-in, streaming and other factors may or may not influence gap detection and continuity judgment in the same way; so a performance score, rather than bias, was used to characterize individual susceptibility to the discontinuity illusion.

Illusory discontinuity and filling-in

There is a strong relationship between the perception of real continuity and perceptual completion of masked sounds, or illusory continuity. With the increase in noise duration, the strength of the continuity illusion increases (χ^2=26.8, N=46, p<0.0001, Friedman ANOVA), closely paralleling the decrease in illusory discontinuity caused by short noises (Figure 5.1). The continuity illusion may not be perceived when the cues supporting the representation of a tone are not fully maintained. This effect was also evident in individual performance: the more reliably an individual judged a continuous tone as continuous (Figure 5.1A), the greater his or her susceptibility was to the continuity illusion (Spearman correlation coefficient ρ=0.87, N=46, p<0.001, Figure 5.5A). The continuity illusion was not perceived by 46% of the participants at 50-ms gap durations (this divides into 28.3% of all participants consistently reported discontinuity, and 17.4% giving responses that were not different from chance; see Table 1). Thirteen percent of the participants never consistently reported hearing the continuity illusion - their performance was either below chance or not significantly different from chance (Figure 5.3C).

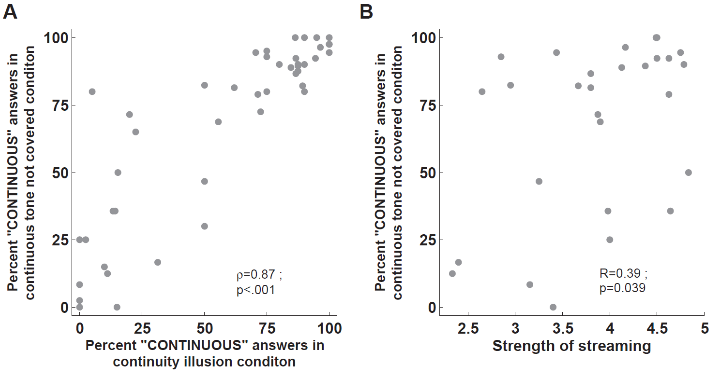


Figure 5.5 A: Relation between perception of illusory and real continuity (data from 50 ms noise duration, N=46) **B:** Streaming ability and subjective disruption of continuity. (N=28). Spearman correlation values are indicated.

Gaps longer than 2000 ms were not used in this experiment, although previous studies (Rieke et al, 2008) and results from pilot experiments (data not shown) suggest that the strength of the continuity illusion declines if longer gaps and noises are used. The present experiment and previous experiments that report relatively long durations of the continuity illusion all used band-pass noises to increase the masking potential of the noise (Riecke et al, 2008, Warren, 1990). Classical stimuli for continuity illusion experiments use white noise, which has a much smaller spectral density and does not completely mask the tone.

Table 5.1 Individual performance in trials with 50ms-long noise.

		Run1	**Run2**	**Total**
Physical continuity	Perform below chance (Hear gaps reliably in a continuous tone, 'Discontinuity illusion')	5	6	11 (24%)
	Perform above chance, (Reliably hear continuous tones as continuous)	12	16	28 (61%)
	Variable performance (not different from chance)	1	6	7 (15%)
Illusory continuity	Perform below chance (do not hear continuity illusion)	6	7	13 (28.3%)
	Reliably perceive continuity illusion	11	14	25 (54.3%)
	Variable performance (not different from chance)	1	7	8 (17.4%)
	Total number of participants	18	28	46

Participants were less likely to detect a gap in a tone just prior to the masking noise as the duration of the noise increased (Figure 5.1D) (Friedman ANOVA, main effect of duration, $\chi^2 = 26.6$, N = 46, p<0.0001). At 50 ms noise durations, all 46 participants performed above the chance level with discontinuous tones, whereas at noise durations of 2000 ms, 15% of the participants consistently reported a discontinuous tone as continuous. Another 17% performed at chance level. This effect may be due in part to a strengthening of backward masking of the gap with an increase in masker duration, which might also act to "fill in" the missing part of the tone (Hashimoto, 2006).

Streaming and illusory discontinuity

The relationship between an individual measure of streaming and susceptibility to disruption of continuity showed that, over all subjects, streaming was positively correlated with susceptibility to the illusory discontinuity of a tone (Spearman ρ = 0.39, N=28, p= 0.039, Figure 5.5B). Participants most susceptible to the 'illusory discontinuity of a tone' were more likely to group the tones in a single stream, whereas participants resistant to the effects of noise were more likely to segregate tones into two distinct streams.

To see how well streaming predicted perception of continuous tone presented with a spectrally remote noise burst condition (Figure 5.1A), a linear discriminant analysis (Matlab Statistics Toolbox, leave-one-out, full cross-validation procedure) was used. In 75% of the cases the classifier was able to correctly determine whether the participant heard continuity reliably or did not in Experiment 1A, solely on the basis of their auditory streaming test data (Permutation test, N = 28, p=0.025).

No relationship between hearing thresholds and continuity/discontinuity perception

No significant relationships were found between tone detection thresholds in noise and continuity/discontinuity perception. There was also no correlation between tone detection thresholds and either the percentage of continuous answers in continuous tone conditions, or for the subjective strength of the continuity illusion (ρ =0.02, p=0.29, and ρ = 0.09, p = 0.65, respectively, all N=18). Kruskal-Wallis ANOVA on data grouped according to the performance in continuous tone conditions using 3 levels (reliably heard continuous tone, chance level performance, reliably heard discontinuity of a tone), also yielded a non-significant effect of hearing threshold (N = 18, p=0.45).

Discussion

One quarter of "normally hearing" individuals consistently reported hearing a continuous tone as interrupted if a noise burst more than an octave away was presented together with the tone. This susceptibility to illusory discontinuity correlates with individual auditory streaming scores; participants that report the tone as interrupted by the noise are more likely to merge the sounds in a single stream. This suggests that such participants loose perceptual integrity of the tone because they group the tone and the noise together into a single perceptual object. Furthermore, the strength of the continuity illusion in an individual subject is strongly dependent on their susceptibility to the discontinuity illusion. Under conditions which perceptually favor the continuity illusion, the perceptual restoration of the tone does not occur as readily in those individuals who are more susceptible to illusory discontinuity.

Variability in individual biases and illusory discontinuity

Participants exhibited notable variability in bias. Bias towards discontinuity reflected an almost perfect performance in detecting discontinuities, but resulted in "mistakes" in continuous tone and illusory continuity perception. Bias towards continuity led to good percent correct scores in the continuous tone condition and strong illusory continuity, but resulted in mistakes in gap detection.

The variability in this bias appears to be the result of individual differences in perception. Many studies have reported large variability in individual susceptibility to non-energetic masking (Neff and Dethlefs, 1995, Lutfi et el, 2003), so the current finding that only 61% of participants perform above a chance level when asked to judge the continuity of a tone in noise is not surprising. What requires explanation is why 24% of the participants consistently judge a continuous tone as discontinuous. This "illusory discontinuity" cannot be attributed to an invariant response bias because the behavior changes with noise duration: in these participants only the short noises can cause illusory discontinuity, while with the longer noises their performance improves and becomes more similar to that of the rest of the participants.

The tone and the noise were separated in frequency by more than an octave, thus the degree of peripheral interaction was minimized[4], following previous research on informational masking (Neff et al, 1987, Lufti et al, 2003, Gutschalk et al, 2008). While other noise frequencies were not used here to test the hypothesis that spectral proximity of the masker defines the strength of illusory discontinuity, previous studies have already demonstrated that target-masker similarity is one of the key factors that influence performance in informational masking paradigms (Durlach et al., 2003, Watson, 1976, Lutfi, 1993; Oh and Lutfi, 1998; Richards and Neff, 2004).

Streaming and illusory discontinuity

The illusory discontinuity of the tone was strongest when the noises were short (50-100 ms). However, participants' performance became progressively better with increasing noise duration. This phenomenon could be due to an increase in subjective frequency resolution with

---

[4] However, with short noise duration it is possible that the spectrum was less defined, and the frequency separation of the tone and the noise was perceptually diminished.

increasing duration of the integration window available for the perceptual judgments, a possibility that needs to be tested in future work.

A direct mechanistic relationship between the processes responsible for streaming and sound perception in noise is supported by the results of Experiment 2. Participants that tended to segregate the tones in two different streams were also more likely to report the tone as continuous. Those participants that grouped the tones in a single stream were more likely to report the tone as discontinuous. Illusory discontinuity might reflect a failure to represent the tone as a separate auditory object, independent from the noise. However, since streaming scores explained only some of the variance in susceptibility to the discontinuity illusion, other factors must play a role. Such factors could include the strength of filling-in, and the ability of a person to consciously report rapidly changing sensory experiences. These and other potential sources of variability need further investigation.

Illusory continuity and discontinuity

If a tone is masked by a brief masking noise, the majority of listeners perceive the tone as continuous even if the obscured portion of the tone has been physically removed from the tone (Miller and Licklider, 1950; Warren, 1972). Such illusorily continuity of tones is thought to be beneficial for auditory perception in complex environments, because it allows the perceptual integrity of sounds to be maintained during intermittent masking (Warren, 1997). Participants that reported a continuous tone as discontinuous in the presence of a spectrally dissimilar masker showed what could be interpreted as an ”impairment” in perceptual restoration. They did not perceive the continuity illusion in trials with short duration, where the continuity illusion has been reported to be at its strongest (Miller and Licklider, 1950; Warren, 1990). However, the same subjects showed a considerable amount of filling-in in trials with longer noise durations (where the same subjects are unlikely to show illusory discontinuity). Filling-in thus appears to happen only under conditions that allow the maintenance of the perceptual integrity of individual auditory objects. This observation is in line with a previous suggestion of Bregman and colleagues that both illusory continuity and streaming depend on a common preliminary mechanism that links together the parts of a sequence that have similar frequencies (Tougas and Bregman, 1990, Bregman et al, 1999).

Familiarization and instructions help to reveal perceptual uncertainty in the continuity illusion

Why did previous studies find that the continuity illusion was strongest in the case of very brief masking noises, whereas in the present study a substantial proportion of the participants perceived the continuity illusion only with longer noises? The difference may be due to the particular instructions and familiarization procedure used in these experiments. Participants were asked to report trials in which there was “a gap or any kind of change in the tone” as discontinuous. Typical instructions in continuity illusion experiments stress only the importance of the gap - “was the tone continuous or did it contain a gap?” A pilot version of the experiment reported here used different instructions (“ignore the noise, attend to the tone, and use the ‘Discontinuous’ button to report the gaps in the tone”). In that pilot experiment, the majority of the participants reported strong continuity illusions with short noise durations, which are consistent with the previous studies (data not shown). The familiarization procedure (without feedback) and the new instructions were added in an effort to better train participants to attend to the tone during the noise and report their percepts.

The experiments described in this chapter found that approximately one quarter of the normally-hearing participants reported a continuous tone as discontinuous in the presence of a short off-frequency noise burst (illusory discontinuity). These participants were also more likely to group sequential tones that differed in frequency together in a simple streaming task, and were less likely to show filling-in under auditory continuity illusion conditions with short noise durations. To test the hypothesis that this individual variation in auditory performance in noise is related to the processing of sound onsets, the experiments reported in Chapter 6 were undertaken: electroencephalographic (EEG) responses were recorded from participants with different perceptual characteristics in response to noise bursts presented alone or together with a tone.

CHAPTER 6. EEG correlates of individual susceptibility to discontinuity illusion

In the experiments reported in the previous chapter, approximately one quarter of the normally-hearing participants consistently reported a continuous tone as interrupted if a noise burst a little more than an octave away was presented together with the tone. One hypothesis for this variation in individual susceptibility to the interrupting effects of remote-frequency noise is that it may be related to individual variation in how onset responses to sounds are processed. If so, differential responses to noise should be evident between listeners who are prone to illusory discontinuity and those who are not even under passive listening conditions. Testing individual performance in passive listening paradigm also assays potential major effects of attention and effort on illusory continuity. Differences between categories of listeners under passive conditions would imply that, while attention and effort may influence perceptual performance of all listeners, these would not explain the differences between listener categories. In the experiment reported in this chapter, EEG responses to noises presented alone or together with tones were recorded while participants focused their attention on a silent movie. The results suggest that a number of attributes describing how individuals differentially parse complex auditory scenes have a statistical relationship with individual differences in two statistically independent attributes of EEG waveform differences at ~50 msec and ~300 msec after noise onset.

Methods

A. Participants

35 subjects (12 males) who were a subset of the participants in the psychophysics experiment described in Chapter 5 were recruited. Their ages ranged from 18 to 32 years (mean ± sd, 23.3 ± 4.3 years). Five of them were left-handed. The study was approved by the Ethics Commission of SISSA. Informed consent for an EEG experiment was obtained from each subject after they were instructed about the nature of the study, in accordance with the principles of the Declaration of Helsinki.

B. Apparatus

The experiments took place in a dimly-lit sound-attenuated room. While a silent movie was played on an LCD monitor, auditory stimuli were presented binaurally through E-A-RTONE 3A insert earphones (Aearo Corporation, Indianapolis, IN). A 128 electrode ActiveOne data acquisition system (Biosemi B.V., Amsterdam, NL, www.biosemi.com) with sintered Ag-AgCl electrodes was used to record EEG signals. Precise synchronization between sound presentation and EEG recording was realized by using a battery-powered audio player: an audio file contained sound stimuli on the track played to the participants, and simultaneous waveform marks on the other track, recorded directly by the EEG hardware and software. EEG data were recorded with a 2048 Hz sampling rate and 16-bit digitization.

C. Stimuli

Two kinds of stimuli were presented: continuous tones accompanied by noises slightly more than an octave above the frequency of the tone (CTNC, Figure 5.1A), and, in separate blocks, the same noise bursts presented alone, without the tone.

In the "tone + noise" blocks stimuli were the same as the continuous tone not covered by noise condition (CTNC) from Experiment 1. The noise (band passed white noise from 2100 to 2700 Hz) appeared 1 second after tone onset and could be either 50 ms or 1000 ms long. These particular durations were chosen because the 50 ms noise burst caused a strong illusory auditory discontinuity, and the longer noise burst made it possible to examine spectrotemporal changes in EEG during the presentation of the noise. The tone was a 1007 Hz sinusoid, either unmodulated or modulated in amplitude by a 40 Hz sinusoid. Intertrial intervals randomly varied from 900 to 1400 ms. Loudness of the tones was set 7 dB below the level of the noises, i.e. at 61 dB HL. The stimuli were presented in pseudorandom order during each 14.5-minute-long recording block.

In the "noise alone" blocks the noise bursts were presented in isolation, without the concurrent tone. All noises were 50 ms long. Two types of noises were presented: centered at 900 Hz or at 2400 Hz; both had a bandwidth of 600 Hz. The two noise types were equiprobable and were presented in pseudorandom order. Intertrial intervals ranged from 900 to 1400 ms. Independently-prepared noise samples were used for each trial. The stimuli were the same for all participants. Stimuli were generated using MATLAB (The MathWorks, Natick, MA) with a 44.1 kHz sampling rate.

D. Procedure

During the EEG experiment participants watched a silent movie of their choice while seated in a comfortable chair 2 meters away from the monitor. They were instructed to concentrate on the movie and neglect the experimental sounds. No responses had to be given. The data were recorded in 2-5 blocks of 10-15 minutes each with pauses in between.

E. Data analysis

ERPs: Preprocessing

Data from 64 channels were resampled at 256 Hz, and band-pass filtered from 1 to 30 Hz using a zero phase-delay filter. An average reference (mean of all 64 channels) was used. The epochs were aligned to the onset of the sound and included 1000 ms pre- and post- stimulus intervals. Two hundred ms of prestimulus time served as a baseline. Epochs that contained movement artifacts were rejected from the analysis, as well as segments that contained eye blinks, using a semiautomatic artifact rejection algorithm. On average, 150 trials in each condition were available from each subject (minimum 63, maximum 303 trials per condition).

ERPs: Statistical procedures

To avoid a multiple comparison problem in the spatial domain, a single frontal-central electrode (Fz in the classical 10-20 coordinate system) was chosen *a-priori* for selection of time intervals of interest. A frontal-central location was chosen for detecting components of evoked response with central topography as well as other components with more frontal topography (Kropotov, 2008). Responses to the CTNC stimulus were examined with respect to individual performance in the psychophysical task.

To quantitatively characterize individual ERP waveforms, the mean voltage of each individual's ERP in an integration window of interest was measured. All the statistical tests used this single measure. Four windows of interest were initially defined by visual data inspection, based on the morphology of waveform peaks (40-66 ms after noise onset; 95-130 ms after noise onset; 160-220 ms after noise onset; 270-350 ms after noise onset). Spearman's correlation coefficient ρ was used to quantify the relationship between ERP and perceptual performance of all 35 participants, including those who heard the CTNC variably (N=9). A non-parametric correlation coefficient was used because the psychophysical performance measure was a ratio that was not distributed normally. P-values of the correlations were corrected for multiple comparisons using the False Discovery Rate correction (Hochberg, Benjamini, 1990).

Non-parametric Schreirer-Ray-Hare 2-way ANOVAs were used to assess the effects of stimulus conditions for the windows of interest in which a significant effect of group was found, and each of the two stimulus comparisons (for a total number of 4 ANOVAs). Reported p-values were Bonferroni-corrected for multiple comparisons.

To verify and better interpret the results of the correlation analyses and ANOVAs with a single recording electrode, the topography of the responses was examined using the data from all recording electrodes. The observation of large amplitude differences distributed in clusters of electrodes between the groups of participants validated that the differences examined here were not artifacts confined to a single electrode. Topographical plots were also used to highlight all electrodes with average voltage values (across subjects) that were correlated with CTNC perceptual scores ($p< .05$, uncorrected for multiple comparisons). This same analysis was also carried out with a stricter threshold ($p<0.005$, see Supplementary Material). Because anatomical MRI data were not available for any of the participants, source localizations were not attempted.

Object-related negativity

We tested a hypothesis discussed previously in the literature that a negative potential found from 100 to 180 ms after stimulus onset may index perceptual segregation of concurrent sounds (Alain, 2007, Zendel and Alain, 2007). These previous studies used mistuned harmonic tones as stimuli, and found that an ERP component peaking at 140 ms after sound onset (based on an integration window from 100 to 180 ms) at fronto-central electrodes distinguished between tuned (heard as a single sound object) and mistuned (heard as multiple sound objects) harmonic sounds. When participants were subsequently asked to report whether they heard the sound as fused into a single complex tone or as two separate sounds, this potential correlated with their perceptual judgments, and was larger when participants perceived the mistuned harmonic as a separate tone (Alain, 2001), leading to the proposal that this object-related negativity (ORN) is a neural indicator of concurrent sound processing.

To test whether concurrent sound-segregation-based cues other than spectral regularity (pitch and timbre) are also reflected in this potential, the responses to noise presented alone and noise presented together with the tone were compared.

Steady State Responses

Trials were also given in which the amplitude of the tone was modulated by a 40 Hz sinusoid with 90% modulation depth. These data underwent the same preprocessing procedures as described above, except that no filtering was applied to them. Power and intertrial coherence at 40 Hz at the Fz electrode was examined in the CTNC condition (Figure 5.1A) during 4 time windows: during tone stimulation before noise onset (0-500 ms after tone onset), during noise stimualtion (0-500 ms and 500-1000 ms after noise onset), and during tone stimulation after the noise offset (1000-1500 ms). Discrete Fourier Transforms (128-point FFTs, $\Delta F=2$ Hz, the epochs tempered with 5 ms cosine square ramps) were used to generate spectra in such a way that 40 Hz was always in the center of a frequency bin. A baseline period of 500 ms before tone onset was also examined. Intertrial coherence, a measure of the reproducibility of the phase of stimulus-locked trial activity at each frequency (Stapells et al., 1984), was calculated using the same number of trials for all participants (63 trials, the minimal number of trials available). Friedman ANOVAs were used to examine the effects of time on power and intertrial coherence. Intertrial coherences were used to examine the relationship between perceptual performance and the steady-state response because they are very robust to interference from noise. Spearman correlations were calculated separately for 3 time windows during the noise and immediately after: 0-500 ms, 500-1000 ms, and 1000-1500 ms after noise onset.

Musical experience

The effects of musical experience were evaluated using data from a questionnaire that participants were asked to complete. They were divided into three groups based on the number of years during which they received musical training (less than 2 years, from 2 to 7 years, more than 7 years).

Results

P50 and $N_{270-350}$ in the tone-plus-noise condition

The relationship between individual susceptibility to continuity disruption and electrophysiological responses to noise were assessed using the same stimuli (the continuous tone not covered by noise (CTNC, Figure 5.1 A)). Significant effects between these variables were found in two time windows out of the four that were analyzed. The figure 6.1 presents grouped data from the participants that exhibited extreme responses, whereas data from all 35 participants were included in the correlation analysis.

Participant less susceptible to continuity disruption by noise showed larger magnitudes of the P50 component of the evoked response (integrated signal voltage from 40 to 66 ms after the onset of the noise); a Spearman correlation of 0.47 (p=0.013, corrected for multiple comparisons) was found at the Fz electrode that was chosen *a-priori* for analysis (Figure 6.1A, 6.2A). Topographic plots revealed that many surrounding electrodes exhibited a similar effect (Figure 6.3A)).

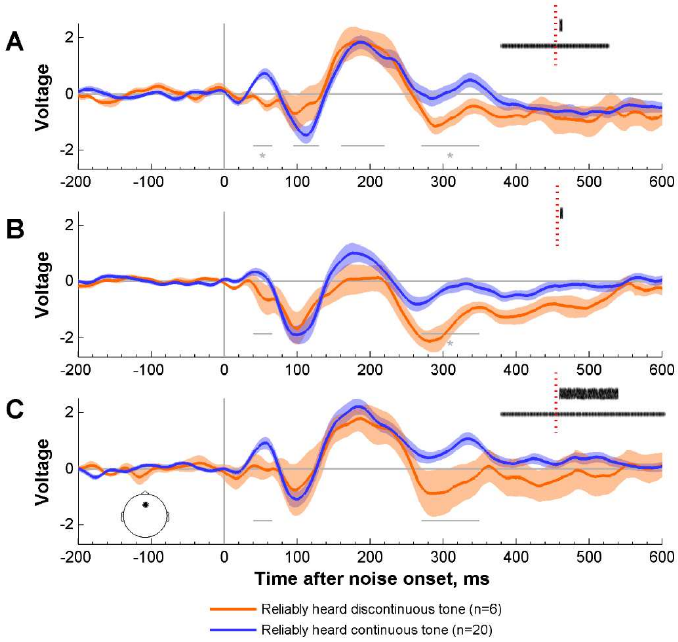


Figure 6.1. ERP aligned to the onset of the noise in 3 conditions (specified also in the upper right angle of the plot): **A.** 50 ms noise burst presented together with the tone **B.** 50 ms noise burst presented without the tone. **C.** 1000 ms noise burst presented together with the tone. Lines are group means and shaded areas are standard errors. Horizontal gray lines denote time windows of interest. Please note that the reported statistical measure was a correlation, which used all 35 subjects (not only the two groups of participants depicted here).

Using an integration time window from 270 to 350 ms after noise onset revealed another EEG waveform feature that showed a significant correlation of integrated signal voltage with perceptual performance (ρ =0.38, p=0.034, corrected for multiple comparisons, Figure 6.2B). The absolute magnitude of the voltage change was greater in the negative direction for the responses of participants who reliably heard the tone as discontinuous; participants who reliably heard the tone as continuous had a similar deflection oriented in the same direction, but of smaller magnitude. Topographically, this effect was also present at multiple electrodes (Figure 6.3B), and appeared to be strongest between 290 and 320 ms. The other two integration windows (105-120 ms, 160-220 ms after noise onset) did not show any significant correlation between signal voltage and behavioral performance (p >0.1).

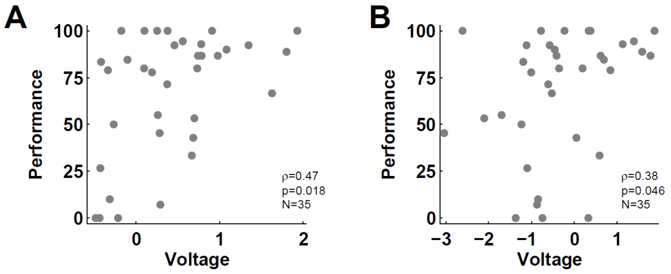


Figure 6.2 Relationship between magnitude of ERP responses and performance when listening to a continuous tone with a spectrally remote 50-ms noise burst **A.** Integrated signal voltage between 40-66 ms after noise onset. **B.** Integrated signal voltage between 270-350 ms after noise onset. Each dot reflects measurement from a single individual. Spearman's ρ and p values (corrected for multiple comparisons) are indicated.

To test for a possible dependence in relationship between the processes indexed by the P50 and the $N_{270\text{-}350}$ components, the correlation between the amplitude of these events was examined. This correlation was not significant or close to significance (ρ=.32, p=.06, uncorrected; the corrected alpha level is 0.0125).

Effects of condition (noise alone vs. tone plus noise)

The observed differences were not confined to the continuous tone plus noise condition, which resulted in the strongest perceptual reporting of discontinuity. Similar effects were also present in responses recorded to the same 50-ms-long noise bursts presented in isolation. The earlier (40-66 ms) time window in the noise alone condition showed a weaker version of the same trend seen above (Figure 6.1 B, 6.3 C). An ANOVA yielded a significant effect of subject performance (p= 0.042, corrected for multiple comparisons), no effect of condition (tone present or not present, p>0.5) and no interaction (p>0.5). However, in this case there was no significant overall correlation between performance and integrated signal voltage at the Fz electrode (ρ=.07, p>0.1).

The 270-350 ms window ANOVA also yielded a significant effect of subject performance (p= 0.02, corrected), no effect of condition (p= 0.31) and no interaction (p>.5). The significant effect of subject performance reflects a similar trend in both conditions: participants that reliably heard the tone as discontinuous had lower values compared to whose that reliably heard the tone as continuous (Figure 6.1 B, Figure 6.2 A). The correlation between performance and voltage was significant at the Fz electrode (ρ=.47, p=.012) in the noise alone condition, and a topographic plot confirmed that this effect was also present at multiple electrode sites (Figure 6.3 D).

Effects of noise duration (50 ms noise vs. 1000 ms noise)

Trials in the tone plus noise condition had noises of two different durations, 50 ms and 1000 ms (Figure 6.1 A and C). The effect of noise duration, assessed by ANOVAs, showed a significant effect of subject performance (p=0.03, corrected) at the earlier (40-66 ms) time window with no effect of noise duration (p>0.5) and no interaction (p>0.5). At the later time window (270-350 ms), the effect of subject performance was also significant (p=0.02), with no significant effect of noise duration (p=0.13) and no interaction (p>0.5). The topography of the differences between the groups of participants was similar in the 1000 ms noise and the 50 ms noise conditions (Figure 6.3 E, F); thus, similar electrophysiological responses are seen at both long and short noise durations.

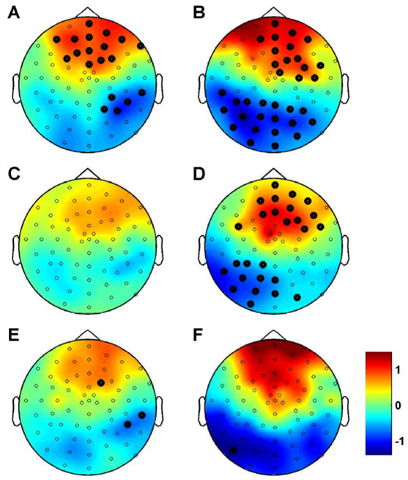


Figure 6.3 Scalp distribution of the potential differences between participants that reliably heard the continuous tone in noise as continuous (N=20) vs. those that reliably heard it as discontinuous (N=6). **Left column** – 40-66 ms integration window; **right column** – 270-250 ms window. **A, B** – 50 ms noise and tone condition; **C, D** – 50 ms noise alone; **E, F** – 1000 ms noise and tone. Red color signifies that the voltage values of the "reliably continuous" group are greater than those of the "reliably discontinuous" group; blue color signifies the opposite (see scale bar for the value of the voltage differences (μV)). Electrodes with nominally significant correlations (uncorrected for multiple comparisons) between performance and voltage (p<0.05) are highlighted.

Relationship between auditory streaming scores and ERPs

The relationship between ERPs and individual auditory streaming scores was examined for the group of 27 participants for whom the sequential auditory streaming scores were available (see Chapter 5, pp. 39 and 46 for details). Neither of the windows of interest yielded a significant correlation [early (40-66 ms after noise onset window): ρ=0.18, p=0.36; late (270-350 ms): ρ = -0.06, p= 0.78; Supplementary Figure S6.4]. This test does not have good power because of the small sample sizes of the two subject types.

Effects of music, age and sex

Effects of musical experience, age and sex on the amplitude of ERPs in the early (40-66 ms after noise onset) and late window of interest (270-350 ms) were not significant (p>0.1 in both cases).

Steady state responses

Both power and intertrial coherence of steady state responses to the amplitude modulation of the tone were significantly affected by the noise burst (p<0.05, corrected): during the noise, the power of the steady state response dropped to 0-5% of its initial value, and intertrial coherence dropped to 15-20% of its initial value. There was no correlation between individual susceptibility to discontinuity and steady state response suppression during the noise (ρ=-0.07, p>0.5) and no correlation between susceptibility to discontinuity and the subsequent recovery of steady state responses after noise offset (ρ=0.02, p>0.5)

Object-related negativity

There was no correlation between hearing the tone continuing under the noise (hearing two auditory objects instead of just hearing the noise burst as a single auditory object) and integrated potential in the window from 100 ms to 180 ms after noise onset (ρ=.05, p>0.5).

Discussion

The current experiment tested the hypothesis that individual susceptibility to illusory discontinuity (subjective disruption of the continuity of a tone in the presence of a spectrally remote noise burst) is related to the neural onset response to the noise burst. These data were collected under the conditions of a passive listening paradigm, in which sounds were presented while participants watched a silent movie of their choice.

Participants less susceptible to illusory disruption of tone continuity by noise showed statistically larger onset responses to noise. This early P50 component of the evoked response to the noise burst (40-66 ms poststimulus) was only present in the group of participants that reliably judged the tone to be continuous in the presence of a noise burst. This finding can be put into context with previous work on the physiological basis of sound source segregation. Previous studies consistently found that sounds belonging to a single stream elicit smaller onset responses when compared with a novel sound or a sound which does not belong to the same stream; the latter events are associated with a larger neuronal onset response in the auditory cortex (Fishman et al, 2001, Micheyl, 2005) and cochlear nucleus (Pressnitzer et al, 2008). In humans, larger electrophysiological onset responses are elicited if participants perceive two melodies at the same time (Gutschalk et al, 2005). Therefore, the “larger” or “more reliably timed” onset responses to noise found in participants who are better able to segregate tones and noises are consistent with both the occurrence of ‘streaming’ (in this case, the segregation of “figure” and “background”) and perhaps a better ability for streaming in subjects who have larger amplitude and/or greater coherence in their brain responses to sound edges. The magnitude of a later ERP component (270-350 ms after noise onset) was larger (that is, greater magnitude of negative voltage) in the participants prone to illusory discontinuity, indicating that the differences are not confined to the initial response to the sound. Both of these findings are discussed in more detail below.

P50

Previous work has suggested two components of the EEG onset response to a sound that have been repeatedly shown to be affected by the organization of sound scenes. P50 (also dubbed P1, seen around 50 ms in the EEG study of Snyder et al (2006), and at 50–70 ms after tone onset in the MEG study of Gutschalk et al (2005)) and N1 (100–120 ms in Snyder et al (2006) and around 100 ms in Guttschalk et al (2005)). Similar patterns of activity have been shown in the primary auditory cortex of awake macaque monkeys: neuronal firing at around 25-70 ms after tone onset is modulated by the properties of the sound that affect source segregation (Micheyl et al, 2005). Novel sounds and sounds which do not belong to a recent sound stream (e.g. the ‘B’ tones in an ABA paradigm) elicit larger onset responses compared to repetitive sounds that form a single stream.

The participants in the present study who were *less* susceptible to disruptive effects of noise showed *larger* onset responses to the noise when it was presented in the context of an ongoing tone “stream”. Indeed, a discernable early P50 ERP component (40-66 ms poststimulus) was present only in the grouped responses of the participants that reliably heard the tone as continuous in the presence of a noise burst. There was a similar trend, but no significant difference in the later N1 component (data not shown). Human intracranial recordings and source analyses of MEG and EEG data have suggested that P50 reflects activity in lateral Heschl’s gyrus and the superior temporal gyrus (Pelizzone et al, 1987, Reite et al, 1988, Liegeois-Chauvel et al., 1994; Gutschalk et al., 2004, Huang et al, 2003). The present data showing that susceptibility to perceptual disruption by noise is reflected in the amplitude of P50 appears consistent with this body of information, and with an FMRI study showing that individual accuracy of sound identification in noise is positively correlated with activation in anterolateral Heschl’s gyrus and the adjacent lateral superior temporal gyrus (Binder et al, 2004).

Both the early (P50) and the later (270-350 ms) components identified here show the same trend in the "noise alone" condition, confirmed by the significant main effect of condition (noise alone/tone plus noise) in the absence of a significant interaction effect. However, the 270-350 ms component in the noise alone condition also showed a significant correlation with performance, suggesting that the observed effect was not specific to the interaction of concurrent sounds. The P50 component did not show a correlation in the noise-alone condition, suggesting that it might be more related to the context–dependent modulation of sound processing. The absence of a significant effect in the noise-alone condition might be due to the fact that, during the "noise alone" recording sessions, similar noises that differed only in their center frequency were presented in an equiprobable fashion in pseudorandom order, and these may have been attributed to the same stream by all listeners. Interestingly, there is no significant correlation between the amplitudes of P_{50} and $N_{270\text{-}350}$ within individual subjects, despite the fact that both of them are significantly correlated with a common performance variable. Unless these results are due to sampling error (based on the number and characteristics of the particular "continuous" and "discontinuous" hearers used in this experiment), then the lack of a correlation suggests that the processes indexed by P_{50} and $N_{270\text{-}350}$ are relatively independent from each other.

$N_{270\text{-}350}$ ("P3a"?)

The magnitude of the negatively-peaked potential between 270 and 350 ms after the onset of the noise was larger in participants susceptible to illusory discontinuity. In the noise alone condition, this negative peak could extend up to 500 ms post stimulus. The topography of the difference wave was fronto-central. It is currently unclear whether this potential has an exact correspondence to other potentials that have been recorded in previous work. In most of the studies that examined ERP correlates of streaming, such late differences (250-600 ms poststimulus) appeared only when the participants attended to the sounds (Hautus and Johnson, 2005, Alain, 2007). In an active oddball paradigm, a positive waveform which appears around 300 ms in response to the attended target stimulus is typically found, a potential called P300 or P3 (Polich, 2007).

In 10-15% of the participants studied by Polich (1988), an oddball sound elicited late responses even without an active task. This component was dubbed ''P3a'' to distinguish it from the task-relevant ''P3b'' potential elicited during target stimulus processing. It was described as a positive-going waveform that appears in response to auditory and visual distracters, and was thought to reflect stimulus-driven activation of prefrontal cortex (reviewed in Polich, 2007). If the late potential difference observed in all the conditions of the present experiment is in fact related to stimulus-driven subliminal attentional processing, then greater "positivity" is observed in the participants who are less susceptible to the effects of noise, suggesting that stronger subliminal attention to the stimuli lead to better sound source segregation. However, the topography of P3a is expected to be central (Polich, 2007), whereas the difference between participants in the present experiment is fronto-central. Therefore, this finding should be interpreted with caution, and further data will be necessary in order to relate this potential to those seen in other work.

Speculations on the origin of intersubject differences

Extrapolating from the relatively small sample of subjects studied here, individual susceptibility to the disruptive effects of noise appears to be continuously distributed in the normal-hearing population, suggesting that a variety of factors may have an influence. Some of these factors may be sensory in nature, and could determine the efficiency of sound encoding (Hornickel et al, 2009). Results of previous studies underscore the importance of stimulus-driven changes in neural representation of sounds, especially stimulus-specific adaptation (Ulanovsky, 2004). The specificity of such adaptation is likely to be a particularly important factor, because it would affect the amount of cross-channel interference and could limit the ability of any putative read-out mechanism to separate neural activity related to different sound sources. Several studies have examined the dependence of the response to a tone on the frequency of the preceding tones (McKenna et al, 1989, Barlett and Wang, 2005, Wehr and Zador, 2005, Eggermont, 2009). In general, preceding tones close to the best frequency of a neuron cause the largest amount of suppression of subsequent responses by that neuron, whereas preceding tones far from the neuron's best frequency may enhance its responses to subsequent tones. This suggests that sharper "adaptation tuning curves" are likely to lead to more efficient segregation of responses to different auditory objects, compared to a wider tuning curves. The time constants of stimulus-specific adaptation are presumably just as important: faster adaptation of the firing rate of non-best frequency neurons could lead to faster segregation of neuronal activity.

Another mechanism is suggested by the literature on modulation of auditory processing by the so-called 'reticular activating system', which is crucial for regulation of sleep-wake states, arousal and attention (reviewed in Reese et al, 1995, Sarter et al, 2005). Early components of auditory responses like P50 are likely to be modulated by this system, which appears to be active during waking and rapid eye movement (REM) sleep, but not during slow wave sleep. The amplitude of P1 is modulated by scopolamine, a cholinergic antagonist (reviewed in Pekkonen, 2005, Buchwald et al, 1991). The suppression of the P1 response to a second click in a sequence of identical clicks is readily observed in EEG recordings from normal human subjects, and the lack of a P1 component in the participants prone to illusory discontinuity might reflect weaker attention modulation that translates into a greater probability of difficulties in perceptually segregating the tone and the noise.

In summary, individual differences in auditory perception in noise appear to be related to variability in early stimulus-driven activation of the auditory cortex, which might be sensory in nature or reflect modulation by subcortical systems. Successful detection and representation of a noise-burst masker appears to favor target-masker segregation and is associated with better auditory object constancy perception in noise.

CHAPTER 7. EEG correlates of the continuity illusion

The aim of this experiment was to investigate neural mechanisms operating during the auditory continuity illusion. This study utilizes the high temporal resolution of EEG to examine the representation of the target sound in the auditory system using a "physically identical stimuli – subjectively different percepts" paradigm. Noise burst loudness was adjusted individually for each subject, in order to have a discontinuous tone covered by a noise burst perceived "correctly" in 50% of the trials as discontinuous, and "incorrectly" in the other 50% of the trials as continuous, i.e. the auditory continuity illusion. Based on previous studies (Sugita, 1998, Petkov et al, 2007), the perception of illusory continuity should be associated with a lack of a subjective "offset" of the illusory tone during the noise, as well as a lack of a subjective "onset" for the tone after the noise is over. This should in turn be associated with reduced or absent neural responses to these events, compared to when the same stimuli are perceived in a "discontinuous" fashion.

However, no evidence was found for any population-level changes in the representation of the tone during the "illusory" and "non-illusory" trials, either in onset/offset responses or in steady state responses (SSRs). In trials that participants identified as continuous (with the continuity illusion) there was, however, lower overall power of beta oscillations (14-35 Hz frequency band, $P<0.001$, uncorrected, $p=.048$, corrected) and a trend for higher power of delta-theta oscillations (4-7 Hz frequency band, $p=0.016$, uncorrected, $p>0.05$, corrected). Most importantly, the oscillatory activity occurring during the baseline period (before the start of the noise period during, which the gap in tone was either correctly detected or perceptually "filled-in") was significantly related to subsequent perceptual judgments. A lower overall amplitude of low-frequency oscillations (<6 Hz frequency band, $p<.02$, corrected) was associated with the subsequent occurrence of the continuity illusion. These findings suggest that the probability of occurrence of the continuity illusion is strongly influenced by changeable characteristics of the pre-existing brain state that the stimuli are interacting with.

Methods:

A. Participants

14 subjects (8 males) were recruited from the participants in the experiment described in Chapter 6 were recruited; only subjects who did not exhibit strong discontinuity bias were allowed to participate. They aged from 18 to 32 years (mean ± sd, 23.5 ± 4.1 years). 12 of them were left-handed. Informed consent for the EEG experiment was obtained from each subject after they were instructed about the nature of the study in accordance with the principles of the Declaration of Helsinki. The study was approved by the Ethics Commission of SISSA.

B. Apparatus

Auditory stimuli were presented binaurally through E-A-RTONE 3A insert earphones (Aearo Corporation, Indianapolis, IN) in a dimly-lit sound-attenuated room. A 128 electrode ActiveOne data acquisition system (Biosemi B.V., Amsterdam, NL, www.biosemi.com) with sintered Ag-AgCl electrodes was used to record EEG signals. Precise synchronization between sound presentation and EEG recording was realized through a battery-powered audio player: an audio file contained sound stimuli on the track played to the participants, and simultaneous marks on the other track, recorded by the EEG hardware and software. The data were recorded with a 2048 Hz sampling rate and 16-bit digitization. Behavioral responses were recorded simultaneously with EEG through special button-boxes, connected to the EEG machine (Biosemi B.V., Amsterdam, NL).

C. Stimuli

In the continuity illusion condition a 2300-ms-long, 1.007 kHz tone contained a gap from 1000 ms to 1500 ms. The gap in the tone coincided with the appearance of a spectrally similar noise (600-Hz-wide band-pass noise, centered at 900 Hz; see Figure 5.1B (p. 35).

On the basis of a psychophysical test conducted prior to the EEG experiment, the loudness of the noise was adjusted separately for each participant to attain 50% continuous and 50% discontinuous judgments for the tone-noise combination. Two noise levels were used for each participant in the continuity illusion condition, usually 1-2 dB apart. On half trials the noise was 2 dB louder than the tone, and on the other half - the ‘continuity illusion’ trials - the noise was 4 dB louder than the tone. This manipulation was done to account for possible variation in thresholds during the experiment, and to insure that performance would be close to 50% for at least some part of the experimental period.

Three control conditions served as references for participants’ perceptual judgments: a continuous tone with a spectrally dissimilar noise (centered at 2400 Hz; bandwidth 600 Hz; see Figure 5.1A, (p. 35), a discontinuous tone with a spectrally dissimilar noise (see Figure 5.1C), and a discontinuous tone accompanied by spectrally similar noise which started 50 ms after the onset of the gap (see Figure 5.1D).

The tones were amplitude modulated at 40 Hz (90% modulation depth). The noise loudness in the reference conditions were set at the mean between the two levels used for the auditory continuity illusion.

The tones were presented at 68 dB hearing level, which none of the participants found to be unpleasant. The stimuli were presented in pseudorandom order during each block. Intertrial intervals varied from 3-4 seconds. Independently-prepared noise samples were used for each trial. The main condition was 8 times more frequent than reference conditions. The stimuli were the same for all participants. Stimuli were generated using MATLAB (The MathWorks, Natick, MA) with a 44.1 kHz sampling rate.

D. Procedure

Participants were instructed to ‘ignore the noise, attend to the tone and judge whether there were any kind of gaps or changes in the tone. A continuous sound should have no such changes, whereas a discontinuous sound could have a gap or a change in any of the properties of the sound during the noise’.

Prior to the EEG experiment, each individual’s threshold for the auditory continuity illusion was determined. The stimuli consisted of a discontinuous tone with spectrally similar noise 500 ms long with a loudness ranging from -4 to 6 dB relative to the tone. Participants were instructed to give their responses by pressing a button that corresponded with their perceptual judgments. The position of the button boxes was switched after each block to avoid lateralization of responses-related potentials. Participants were instructed to respond after the end of the second part of the tone. To avoid discomfort associated with maintaining visual fixation during the experiments, the participants were instructed to keep their eyes closed during the experiment. The data were recorded in 3-5 blocks of 10 minutes each with pauses in between.

E. Analysis

Behaviour

Reaction times were analyzed within individual subjects using nonparametric Scheirer-Ray-Hare ANOVAs with two main factors: participants' judgements (2 levels-"Continuous" (continuity illusion) or 'Discontinuous' (the gap was correctly detected)) and noise loudness (2 levels - louder noises and softer noises). For group analyses, individual subjects' medians in the four conditions were submitted to Scheirer-Ray-Hare ANOVA using the same main factors.

ERPs

Data from 64 channels were resampled at 256 Hz, and band-pass filtered from 1 to 30 Hz using a zero phase-delay filter. Average reference was used. The epochs included 1000 ms pre- and 500 ms post- stimulus intervals. Epochs that contained movement artifacts were rejected, as well as segments that contained eye blinks, using a semiautomatic artifact rejection algorithm. On average, 83 trials in each condition were available from each subject (minimum 63, maximum 303 trials per condition).

EEG responses to the continuity illusion stimuli were examined at the *a-priori* chosen Fz electrode (as in the experiment in Chapter 6). Since an effect was predicted to be present at the time of noise onset and after the noise offset, analyses were confined to the time intervals from the onset of the noise to 500 ms after the noise offset. A baseline was used from 200 ms before the noise onset to the noise onset (700-1000 ms after the onset of the tone). Sign-tests for dependent samples were used to assess the effects of a participant’s perceptual judgment in the time windows of interest, which were defined by the peaks of the mean EEG signal.

Steady state response (SSR)

For time-frequency analysis the data underwent the same procedures as described above except that no filtering was applied to the data.

The power and the phase of responses at 40 Hz (the frequency of the amplitude modulation) was calculated in three time windows: during the first part of the amplitude modulated tone (0-500 ms after the onset of the tone), during the noise (1000-1500 ms after the tone onset), and after the noise offset by means of Discrete Fourier transforms, the output of which was centered at the target frequency of 40 Hz (128-point FFTs, $\Delta F=2$ Hz, the epochs tempered with 5 ms cosine square ramps).

In the trials in which the tone was judged as continuous, the steady state response was further assessed by means of Rayleigh's test for nonuniformity of the phase distribution across trials (Toolbox for circular statistics with Matlab, Philipp Berens). The test determined whether there is a consistent clustering of the phase of the EEG signal at 40 Hz. It was performed on each participant’s data individually. During the tone (0-500 ms after the tone onset) all 14 participants showed significant phase clustering ($p \ll .001$) that assures good quality of the SSR signal. The SSR group analysis was performed using a dependent samples sign test on power and intertrial coherence measures. The null hypothesis of this test was that there was no difference in power or intertrial coherence between the baseline period (from -500 to 0 ms before the tone onset) and the noise period (from 1000 to 1500 ms after the tone onset) in the trials on which the tone was judged as continuous.

Time-frequency analysis

The oscillatory power was calculated using Discrete Fourier transforms (FFT size = 128 points, $\Delta F=2$ Hz, the epochs tempered with 5 ms cosine square ramps) in two 500-ms long time windows of interest: during the noise (from 0 to 500 ms after the onset of the noise, 1000 msec after tone onset), and after the noise offset (1500-2300 ms after tone onset). Statistical analyses were performed using the Fieldtrip toolbox for EEG/MEG-analysis (Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, the Netherlands; http://www.ru.nl/neuroimaging/fieldtrip). A paired-sample permutation t-test with cluster-based correction for multiple comparisons (Maris and Oostenveld, 2007) was performed on the spectra integrated within a band of interest.

Since the cluster-based correction method for non-parametric ANOVA was problematic to implement, noise loudness effects (two levels within each subject - softer and louder noises, 1-2 dB apart) were analyzed using the means of the average spectra of the trials with different noise levels for each subject. Reported p-values were corrected for multiple comparisons (n=8) using the Benjamini - Hochberg false discovery rate; uncorrected cluster-level p-values were also examined.

No baseline-correction for power values was used in these analyses (see the next paragraph below).

To visualize steady-state response (SSR) time courses, time-frequency transforms were calculated using Morlet wavelets (EEGlab toolbox, Delorme and Makeig, 2004). The plotted values represent the power with respect to the baseline (the silent 500 ms period before the tone onset), which was calculated using all the trials of a given participant and thus is common for the two conditions.

Analysis of the pre-noise period

Physically identical continuity illusion stimuli (discontinuous tone plus the spectrally similar noise) can elicit either continuous or discontinuous responses when the noise loudness is set in the ambiguous region. The ability to attend to the stimuli may vary during the experiment. Cyclic alternations of behavioral states have been shown to occur in low-arousal experimental environments (Conte et al, 1995, Makeig and Jung, 1995). Thus, the brain state during which the trial is presented may favor either the continuity illusion or gap detection. The power of neural oscillatory activity during the tone but before the appearance of the noise was examined to see whether this bore any relation to the probability of elicitation of the continuity illusion during the immediately following noise-burst period. The following frequency bands were examined: delta (from 1 to 4 Hz), theta (4- 7 Hz), alpha (8-12 Hz), beta (13-35 Hz) and gamma (35-80 Hz). The temporal integration window spanned 500 ms of the silent intertrial interval and the first 500 ms after tone onset. Paired-sample permutation t-tests with cluster-based correction for multiple comparisons (Maris and Oostenveld, 2007) were performed on the spectra integrated within a band of interest. The Benjamini - Hochberg false discovery rate was used to correct the p-values for multiple comparisons (n=5).

Results

Behavior

Louder noises were associated with a higher probability of the continuity illusion: there was a significant effect of noise loudness on perceptual judgment in the “discontinuous tone presented together with spectrally similar noise” condition ($p<.01$, N=14, Friedman ANOVA on the scores obtained in the psychophysical test done prior to the EEG experiment): louder noises were associated with a higher number of 'Continuous' responses. The mean loudness level of the noises in the EEG experiment was 1.25 dB higher than the tone (standard deviation = 1.9, n=12 (for 2 participants the loudness values were not registered), it ranged from -1 dB (average of the two noise levels used for the participant, which were -2 and 0 dB compared to the loudness of the tone) to 6.5 dB (see the distribution of the mean noise loudness in Supplementary Figure S7.1).

A group analysis of the effect of perceptual judgment on reaction times was not significant (n=14, p>0.5, Scheirer-Ray-Hare Two-way ANOVA).

ERPs

During the noise and after the noise offset there were no differences between the trials on which the tone was judged as continuous compared to the trials on which the gap was detected (n = 14, p>.05, paired-samples t-test with cluster-based correction for multiple comparisons).

Steady state response (SSR)

The steady-state response signal at the frequency of amplitude modulation of the tone (40 Hz) is observable during the presentation of the tone (Figure 7.1, first portion of the tone (between 0 and 1000 ms after the tone onset) and the second portion after the noise offset (between 1500 and 2300 ms)).

Only 1 out of the 14 participants showed any suggestion of clustering of the phase at 40 Hz during the noise in the trials judged as continuous, but this effect did not survive multiple comparison correction (p=0.038, uncorrected, the multiple comparison correction threshold in this case is 0.0018, Rayleigh's test for nonuniformity of the phase distribution).

There were no differences in power and intertrial coherence at 40 Hz between the baseline period and the noise period (from 1000 to 1500 ms after tone onset) in the trials on which the tone was judged as continuous and those on which it was judged as discontinuous (n=14, p<0.1, sign test).

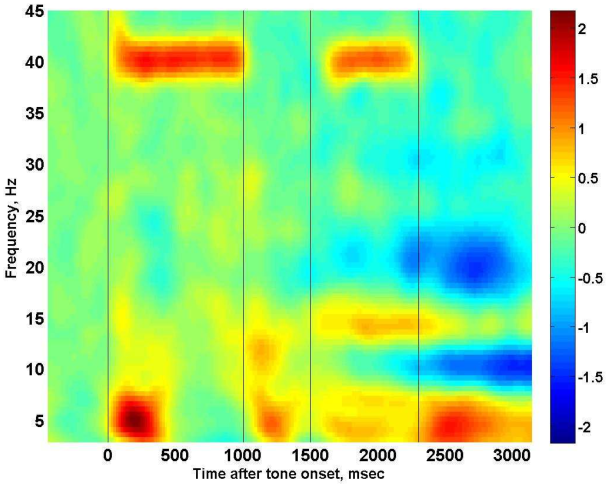


Figure 7.1 Event-related spectral perturbations at the Fz electrode (all trials pooled together irrespective of perceptual judgment). The black lines denote (from left to right) the onset of the tone, the onset of the noise, the offset of the noise, and the offset of the tone. There is no tone present during the noise. The signal at 40 Hz reflects the amplitude-modulation of the tone (steady-state response, SSR).

Oscillatory activity during the noise

The relationship between EEG oscillatory during the noise and the reappearance of the tone was examined for a correlation with perceptual judgments.

During the continuity illusion-inducing noise, differences are present in the beta frequency range (14-35 Hz). The trials on which the tone was judged as continuing behind the noise had less power in the beta band when compared to the trials on which the gap was correctly detected (p=0.001, uncorrected; p=.048, corrected). In the theta frequency band (4-7 Hz), power was greater in the trials judged as continuous (p=0.015 uncorrected, p<0.05, corrected). Table 2 presents the analysis of power in four frequency bands (cluster-based permutation paired t-test, N=14).

	Theta (4-7 Hz)	Alpha (8-12 Hz)	Beta (14-35 Hz)	Gamma (35-80 Hz)
During the noise (0 to 0.5 s after noise onset)	p=0.016 (uncorrected); c.s. -20.9 C>D	n.s.	P<0.001 (uncorrected);; c.s. 33.5D>C	n.s.
After noise offset / onset of the second portion of the tone	n.s.	n.s.	n.s.	n.s.

Table 2 Results of the time-frequency analysis. Dependent samples t-tests between average power in trials judged as continuous and as discontinuous (N=14); n.s. stands for p>0.05, c.s. = cluster statistic; C>D = power in trials judged as continuous is significantly higher than in trials judged as discontinuous.; D>C= power in trials judged as discontinuous is significantly higher than in trials judged as continuous. The time windows in all cases span 500 ms.

Oscillatory activity during the baseline period

Physically identical continuity illusion stimuli (a discontinuous tone plus spectrally similar noise) elicited either continuous or discontinuous responses when the noise loudness was set in the ambiguous region. To see whether perceptual judgments were related to neural activity during the brain state immediately prior to the noise presentation, the relationship between oscillatory activity in the pre-noise period and the subsequent perceptual judgments was examined.

There was a significant difference between the trials subsequently judged as continuous and as discontinuous in the power of delta (p=0.0030 (uncorrected; p=0.0150 (FDA-corrected)) cluster-based statistic 33.4892, N=14) and theta (p=0.0080 (uncorrected; p=0.020 (FDA-corrected)) cluster-based statistic 28.9585, N=14). There were no effects in the alpha, beta, or gamma frequency ranges.

Given the close proximity of the delta and theta frequency bands and the same direction of the effect, it appears that greater power of low-frequency oscillations (<6 Hz) is present during the pre-noise period in the trials subsequently judged as discontinuous (figure 7.2).

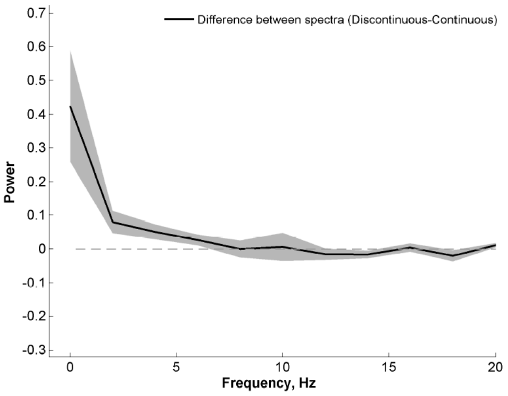


Figure 7.2 Difference between the energy spectra of trials judged as discontinuous and continuous in the pre-noise period (which included both 500 ms of silence and the first 500 ms of the tone). The line and the shaded area represent the mean and the standard error of the individual difference waves (N=14). The data shown are the mean across all 64 channels.

Discussion

To investigate mechanisms of illusory continuity, participants were engaged in an active task in which they were instructed to detect gaps or any kind of changes in target tones while ignoring any noise bursts that might have occluded a gap. The loudness of the noise was adjusted individually so that physically identical stimuli elicited both 'continuous' and 'discontinuous' judgments. We examined responses to tones and noises on the trials in which the gap in the tone was correctly detected, and the trials on which the tone appeared intact (the auditory continuity illusion).

EEG correlates of the auditory continuity illusion

Illusory continuous tones are represented in the auditory cortex similarly to physically continuous tones: sustained-responder neurons fire throughout the noise, and onset- and offset-responders do not change their firing rate at the edges of the gap filled by the noise (Petkov et al, 2007). This previous study, however, used *physically dissimilar* sounds that elicited strong continuity illusions or favored gap detection, thus making it difficult to determine the exact nature of neural firing differences that are necessary and sufficient for the perception of illusory continuity, which is a key question that must be answered in order to understand the neural mechanisms associated with the continuity illusion.

In the present study, the loudness of the noise was adjusted so that tones in the presence of physically identical noise could be judged as either continuous or discontinuous. Discontinuous judgments corresponded to the correct detection of the gap in the tone, while the trials on which the tone was perceived as continuous presented the case of illusory continuity. Larger offset responses at the “end” of the tone/beginning of the noise were expected in the trials on which the tone was judged as discontinuous. However, we found no significant effect of perceptual judgment on the magnitude of the evoked potentials at the onsets or offsets of the noise. This observation is consistent with a similar study which also failed to find a significant effect of perceptual judgment on the magnitude of scalp-recorded event related potentials to physically identical stimuli (Riecke et al, unpublished).

There may be several possible reasons why these EEG studies have failed to find this effect. First, ERP activity after noise onset may reflect an onset response to the (louder) noise to a greater extent than an offset response to the (weaker) tone, and noise stimuli probably engage more spatially-extended tonotopic neuronal responses than tonal stimuli. The number of trials (approximately 80 per condition in both experiments) may not have been sufficient to detect weak offset responses to the tone in noise. Second, evoked responses may be altered by the presence of concurrent noises (Ozdamar and Bohorquez, 2008). For example, the timing of the offset response in noise may become more inconsistent across trials, leading to an impoverished ability to detect them in averaged EEG signals. Additionally, the strength of a cortical offset response may be substantially diminished during noise. These issues, which are potentially very important for characterizing auditory behavior in noisy environments, might be resolved by means of animal electrophysiological studies in which stimulus-specific neural activity could be recorded during perception of ambiguous stimuli in the presence of noise.

Occurrence of the continuity illusion was accompanied by a significant increase in beta-band power (14-35 HZ) and decrease in delta-theta band (4-7 Hz) oscillatory activity during the noise. The changes in power in the beta band have been previously associated with attentive sensory processing and sensory-motor coordination (Baker, 2007), as well as with gating of sensory responses in a paired stimulus paradigm (Hong et al, 2008). EEG-recorded delta-theta activity has been associated with the synchronization of neuronal assemblies in cognitive functions; its power appears to increase with working memory load and mental effort (Mizuhara et al, 2004).

Differences in the EEG spectrum immediately prior to the presentation of the continuity illusion-inducing noise were related to the subsequent perceptual judgment. The amplitude of low-frequency oscillations (<6 Hz) was higher on trials in which the tones were judged as discontinuous. Such an effect could conceivably be related to a subject’s degree of wakefulness or selective, conscious attention. We conclude that the probability of eliciting the auditory continuity illusion is affected by a listener’s brain state. This finding represents a challenge for the interpretation of data obtained in ‘same stimuli-different perceptions’ experiments, because the activity associated with one or another perceptual judgment might actually be determined by the differences in pre-existing brain state rather than by perceptual processing itself.

Selective attention has been shown to increase the probability of perceiving filling-in compared to unattended figures in the visual domain (De Weerd, 2006). This effect was augmented by boosting attention. In the auditory domain, similar enhancement effects of attention for the continuity illusion will be difficult to assess, because increasing attention may also facilitate gap detection. From one point of view, increasing the power of low-frequency oscillations may be interpreted as evidence for decreases in alertness and/or drowsiness (Makeig and Jung, 1995). If so, the occurrence of the auditory continuity illusion in trials with relatively low power of low-frequency oscillation could be evidence for a lack of attention favoring perceptual restoration. Clearly, the present experiments suggest the need for more research exploring the relationship between auditory filling-in and attention.

Illusory continuity of steady amplitude-modulated sounds

Tones are judged by participants as "unchanged" during the auditory continuity illusion. The frequency of an amplitude modulation introduced into the tone is closely matched between the physically present and illusory portions of the tone (Lyzenga et al, 2005). The portion of the present experiment that utilized the steady-state response (SSR) could be regarded as a test of the hypothesis that the illusory continuity of steady amplitude modulated sounds depends upon the persistence of rhythmic activity triggered by synchronization of neural responses to the sound envelope. During the portion of the tone that preceded and followed the noise burst, there was an extremely robust rhythmic oscillation of the EEG signal (as reflected by strong inter-trial coherence) at the frequency of the amplitude modulated sound, with a very good signal-to-noise ratio. During the noise after an amplitude-modulated tone, no rhythmic activity was observed at the 40-hz oscillation rate during the continuity illusion, yet participants still reported hearing the tone unchanged on those trials. Furthermore, the SSR for an amplitude-modulated tone was also drastically reduced in the presence of a spectrally remote noise which does not mask the tone (see Chapter 6, p. 57).

Both the phase and the amplitude of the steady-state response have previously been shown to be perturbed by a noise click presented to the ear contralateral to the one in which an amplitude-modulated tone is presented (Ross et al, 2005). Cortical P1 (P50) to speech or noise is significantly affected by a stimulus presented to the contralateral ear, but earlier response components (such as the ABR and Pa) are not (Ozdamar and Bohorquez, 2008). These data may point to inhibitory mechanisms which prevent the synchronization of large neuronal populations to the envelope of the incoming sound. Similar phenomena have been observed in studies of comodulation masking release[5] in the cochlear nucleus and the auditory cortex. Addition of a comodulated flanker tone partially suppresses locking of spike activity in the cochlear nucleus to

---

[5] If a pure tone is added to the dips of a sinusoidally amplitude-modulated pure tone masker, the target tone is detected better in the presence of a flanker tone in a spectrally distant region whose amplitude is comodulated with the masker tone.

the envelope of the masker, thus facilitating the detection of the target (Pressnitzer et al, 2001, Winter et al 2005). In the majority of the neurons in cat auditory cortex, addition of a spectrally remote pure tone abolished the phase-locking of the membrane potential to the modulated noise (Las et al, 2003). Thus, active mechanisms might be acting in the auditory system that favor cross-channel neuronal integration of auditory information in some circumstances and desynchronize processing in the presence of a noise.

With respect to the neural mechanisms of the auditory continuity illusion, the lack of any large-scale entrainment in oscillatory activity during the continuity illusion with amplitude-modulated tones has now been found independently in two experiments (the present one and Pressnitzer et al., unpublished). This does not exclude the possibility that a small subset of neurons may maintain the rhythm of the amplitude-modulated sound. However, these results are in line with the psychophysical observation that the phases of an amplitude-modulated or frequency-modulated sound are not perceptually preserved during the continuity illusion (since participants cannot discriminate between sounds in which the portion of the tone after the noise is either shifted by 90 degrees or is kept in accordance with the first portion of the tone (Lyzenga et al., 2005)). In addition, neurometric thresholds for auditory perceptual flutter based on the periodicity of firing in primary auditory cortex are lower than any individual monkey’s behavioral thresholds (Lemus et al, 2009), while firing-rate-based neurometric thresholds were very similar to behavioral discrimination thresholds. This may suggest that rate-based coding is more likely to be involved in the generation of the continuity illusion than fine-scale temporal codes that would need to be maintained during the illusion period.

Conclusions and future directions

Illusory discontinuity is likely to be a form of informational masking since it appears also when the noise is spectrally dissimilar (slightly more than one octave away). It is suggested that this arises due to grouping of the noise and the tone in a single auditory object, in which the perceptual identity of the tone is lost. Surprisingly, participants who are most resistant to the effects of noise (i.e. consistently perceive the tone as continuous) show larger onset responses for the noise, a finding consistent with a streaming hypothesis: responses to sounds that belong to the same stream progressively habituate, whereas sounds that belong to a new stream elicit larger onset responses (Fishman et al, 2001, Gutschalk et al, 2005).

The continuity illusion is dependent on the preservation of the identity of the target sound during the noise, thus it is tightly related to streaming (Bregman, 1990). It was shown here that participants that cannot maintain perceptual integrity of a continuous sound in noise also are much less likely to experience the continuity illusion, especially during short noise bursts which have previously been shown to elicit the strongest continuity illusions (Warren, 1999). With increasing noise duration, however, the probability of experiencing the continuity illusion grows stronger in these participants, paralleling the decrease in strength of illusory discontinuity.

It is assumed that illusory discontinuity reported in the lab parallels difficulties with auditory perception experienced in real, complex environments. The quantification of this relationship is an important future direction for this work. One could expect that a correlation between these two phenomena would be rather strong, as anecdotally one of our participants with a pronounced discontinuity illusion once went to an otolaryngologist to ask whether something could be done with his inability to understand speech in noise. The doctor found no hearing deficit and thus had no suggestions.

It is worth pointing out some of the interesting directions for future research on this topic. If illusory discontinuity is primarily associated with difficulties for auditory perception in the presence of noise, the number of individuals that exhibit this effect (approximately 24% of the small study population of 46 people examined here) seems rather alarming. Could there be any beneficial effects for sensory processing associated with this characteristic? Might there also be a non-random distribution of psychological profiles associated with susceptibility to illusory discontinuity? Can training affect the susceptibility to illusory discontinuity? How might illusory discontinuity be modulated by pharmacological agents? It is ultimately hoped that a better understanding of the neuronal contributions to the effects studied in this thesis will eventually lead to therapies that could help individuals cope with auditory noise interference, especially ageing individuals, many of whom specifically complain about their problems with speech comprehension in noise.

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, Ogawa KH, Woods DL (1996) Aging and the segregation of auditory stimulus sequences. J Gerontol B Psychol Sci Soc Sci 51:P91–P93

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

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

Alain C, Quan J, McDonald K, Roon PV (2009) Noise-induced increase in human auditory evoked neuromagnetic fields. Eur J Neurosci 30:132–142

Alain C, Reinke K, He Y, Wang C, Lobaugh N (2005) Hearing two things at once: neurophysiological indices of speech segregation and identification. J Cogn Neurosci 17:811–818

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

Alain C, Snyder JS (2008) Age-related differences in auditory evoked responses during rapid perceptual learning. Clin Neurophysiol 119:356–366

Alain C, Snyder JS, He Y, Reinke KS (2007) Changes in auditory cortex parallel rapid perceptual learning. Cereb Cortex 17:1074–1084

Allen P, Wightman F (1992) Spectral pattern discrimination by children. J Speech Hear Res 35:222–233

Allen P, Wightman F, Kistler D, Dolan T (1989) Frequency resolution in children. J Speech Hear Res 32:317–322

Anokhin AP, Vedeniapin AB, Heath AC, Korzyukov O, Boutros NN (2007) Genetic and environmental influences on sensory gating of mid-latency auditory evoked responses: a twin study. Schizophr Res 89:312–319

Baker SN (2007) Oscillatory interactions between sensorimotor cortex and the periphery. Curr Opin Neurobiol 17:649–655

Baldeweg T (2006) Repetition effects to sounds: evidence for predictive coding in the auditory system. Trends Cogn Sci 10:93–94

Baldeweg T, Klugman A, Gruzelier J, Hirsch SR (2004) Mismatch negativity potentials and cognitive impairment in schizophrenia. Schizophr Res 69:203–217

Bartlett EL, Wang X (2005) Long-lasting modulation by stimulus context in primate auditory cortex. J Neurophysiol 94:83–104

Bar-Yosef O, Nelken I (2007) The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. Front Comput Neurosci 1:3

Bar-Yosef O, Rotman Y, Nelken I (2002) Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J Neurosci 22:8619–8632

Beauvois MW, Meddis R (1996) Computer simulation of auditory stream segregation in alternating-tone sequences. J Acoust Soc Am 99:2270–2280

Bee MA, Klump GM (2004) Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J Neurophysiol 92:1088–1104

Berglund B, Hassmen P, Job RF (1996) Sources and effects of low-frequency noise. J Acoust Soc Am 99:2985–3002

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295–301

Boer JD, Thornton ARD (2008) Neural correlates of perceptual learning in the auditory brainstem: efferent activity predicts and reflects improvement at a speech-in-noise discrimination task. J Neurosci 28:4929–4937

Bonino AY, Leibold LJ (2008) The effect of signal-temporal uncertainty on detection in bursts of noise or a random-frequency complex. J Acoust Soc Am 124:EL321–EL327

Boutros NN, Belger A (1999) Midlatency evoked potentials attenuation and augmentation reflect different aspects of sensory gating. Biol Psychiatry 45:917–922

Braaten, Reynolds (1999) Auditory preference for conspecific song in isolation-reared zebra finches. Anim Behav 58:105–111

Bradlow AR, Kraus N, Hayes E (2003) Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res 46:80–97

Bregman AS, Ahad PA, Crum PA, O'Reilly J (2000) Effects of time intervals and tone durations on auditory stream segregation. Percept Psychophys 62:626–636

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

Bregman AS, Dannenbring GL (1977) Auditory continuity and amplitude edges. Can J Psychol 31:151–159

Bregman A (1990) Auditory scene analysis. MIT, Cambridge, MA.

Brosch M, Schreiner CE (1997) Time course of forward masking tuning curves in cat primary auditory cortex. J Neurophysiol 77:923–943

Brungart DS (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. J Acoust Soc Am 109:1101–1109

Buchwald J, Rubinstein E, Schwafel J, Strandburg R (1991) Midlatency auditory evoked responses: differential effects of a cholinergic agonist and antagonist. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 80:303 - 309

Buonomano DV, Maass W (2009) State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci 10:113–125

Calford MB, Semple MN (1995) Monaural inhibition in cat auditory cortex. J Neurophysiol 73:1876–1891

Carandini M, Ferster D (1997) A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. Science 276:949–952

Carlyon R. P. JMD (2009) Changes in the Perceived Duration of a Narrowband Sound Induced by a Preceding Stimulus. Journal of Experimental Psychology: Human Perception and Performance n/d:n/d

Chait M, Poeppel D, Simon JZ (2006) Neural response correlates of detection of monaurally and binaurally created pitches in humans. Cereb Cortex 16:835–848

Chalikia MH, Bregman AS (1989) The perceptual segregation of simultaneous auditory signals: pulse train segregation and vowel segregation. Percept Psychophys 46:487–496

Ciocca V, Bregman AS (1987) Perceived continuity of gliding and steady-state tones through interrupting noise. Percept Psychophys 42:476–484

Conte S, Ferlazzo F, D'Olimpio F (1995) Rhythmic variations in reaction times and evoked potentials. Int J Neurosci 82:127–133

Crum PAC, Hafter ER (2008) Predicting the path of a changing sound: velocity tracking and auditory continuity. J Acoust Soc Am 124:1116–1129

Cusack R, Roberts B (1999) Effects of similarity in bandwidth on the auditory sequential streaming of two-tone complexes. Perception 28:1281–1289

Cusack R, Roberts B (2000) Effects of differences in timbre on sequential grouping. Percept Psychophys 62:1112–1120

Cusack R, Roberts B (2004) Effects of differences in the pattern of amplitude envelopes across harmonics on auditory stream segregation. Hear Res 193:95–104

Dannenbring GL, Bregman AS (1976) Stream segregation and the illusion of overlap. J Exp Psychol Hum Percept Perform 2:544–555

Darwin C (2005) Pitch and Auditory grouping, in Pitch: Neural coding and Perception P. E. al, ed. Springer.

Dau T, Verhey J, Kohlrausch A (1999) Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. J Acoust Soc Am 106:2752–2760

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

Denham (2001) Cortical synaptic depression in auditory perception in Computational models of auditory function S. Greenberg, ed.

Denham SL, Winkler I (2006) The role of predictive models in the formation of auditory streams. J Physiol Paris 100:154–170

Deutsch D (1974) An auditory illusion. Nature 251:307–309

Deutsch D (1975) Musical illusions. Sci Am 233:92–8,103-4

Diaz-Quesada M, Maravall M (2008) Intrinsic mechanisms for adaptive gain rescaling in barrel cortex. J Neurosci 28:696–710

Drake C, McAdams S (1999) The auditory continuity phenomenon: role of temporal sequence structure. J Acoust Soc Am 106:3529–3538

Dreschler WA, Plomp R (1985) Relations between psychophysical data and speech perception for hearing-impaired subjects. II. J Acoust Soc Am 78:1261–1270

Duifhuis H, Willems LF, Sluyter RJ (1982) Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception. J Acoust Soc Am 71:1568–1580

Durlach NI, Mason CR, Shinn-Cunningham BG, Arbogast TL, Colburn HS, Kidd G (2003) Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. J Acoust Soc Am 114:368–379

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

Dyson BJ, Alain C (2008) It all sounds the same to me: sequential ERP and behavioral effects during pitch and harmonicity judgments. Cogn Affect Behav Neurosci 8:329–343

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

Elhilali M, Shamma SA (2008) A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. J Acoust Soc Am 124:3751–3771

Ermutlu MN, Demiralp T, Karamursel S (2007) The effects of interstimulus interval on sensory gating and on preattentive auditory memory in the oddball paradigm. Can magnitude of the sensory gating affect preattentive auditory comparison process? Neurosci Lett 412:1–5

Erwin RJ, Buchwald JS (1986) Midlatency auditory evoked responses: differential recovery cycle characteristics. Electroencephalogr Clin Neurophysiol 64:417–423

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

Fletcher H (1940) Auditory patterns. Rev Mod Phys 12:47-65

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

Gabrieli JD (1998) Cognitive neuroscience of human memory. Annu Rev Psychol 49:87–115

Galambos R, Makeig S (1992) Physiological studies of central masking in man. II: Tonepip SSRs and the masking level difference. J Acoust Soc Am 92:2691–2697

Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci U S A 78:2643–2647

Garcia-Rill E, Moran K, Garcia J, Findley WM, Walton K, Strotman B, Llinas RR (2008) Magnetic sources of the M50 response are localized to frontal cortex. Clin Neurophysiol 119:388–398

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: a review of underlying mechanisms. Clin Neurophysiol 120:453–463

George ELJ, Festen JM, Houtgast T (2006) Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. J Acoust Soc Am 120:2295–2311

Grataloup C, Hoen M, Veuillet E, Collet L, Pellegrino F, Meunier F (2009) Speech restoration: an interactive process. J Speech Lang Hear Res 52:827–838

Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc Natl Acad Sci U S A 100:253–258

Grier JB (1971) Nonparametric indexes for sensitivity and bias: computing formulas. Psychol Bull 75:424–429

Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23

Grimault N, Micheyl C, Carlyon RP, Arthaud P, Collet L (2000) Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. J Acoust Soc Am 108:263–271

Grimault N, Micheyl C, Carlyon RP, Arthaud P, Collet L (2001) Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing. Br J Audiol 35:173–182

Grimault N, Bacon SP, Micheyl C (2002) Auditory stream segregation on the basis of amplitude-modulation rate. J Acoust Soc Am 111:1340–1348

Grossberg S, Govindarajan KK, Wyse LL, Cohen MA (2004) ARTSTREAM: a neural network model of auditory scene analysis and source segregation. Neural Netw 17:511–536

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

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

Gutschalk A, Oxenham AJ, Micheyl C, Wilson EC, Melcher JR (2007) Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. J Neurosci 27:13074–13081

Haenschel C, Vernon DJ, Dwivedi P, Gruzelier JH, Baldeweg T (2005) Event-related brain potential correlates of human auditory sensory memory-trace formation. J Neurosci 25:10494–10501

Hall JW, Peters RW (1981) Pitch for nonsimultaneous successive harmonics in quiet and noise. J Acoust Soc Am 69:509–513

Hartmann, WM (1998) Signals, Sounds, and Sensation. New York: Springer-Verlag.

Hashimoto A, Watanabe S, Inui K, Hoshiyama M, Murase S, Kakigi R (2006) Backward-masking: the effect of the duration of the second stimulus on recognition of the first stimulus. Neuroscience 137:1427–1437

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

Heil P (2001) Representation of sound onsets in the auditory system. Audiol Neurootol 6:167–172

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

Heinrichs RW (2004) Meta-analysis and the science of schizophrenia: variant evidence or evidence of variants? Neurosci Biobehav Rev 28:379–394

Henderson D, Salvi RJ (1998) Effects on noise exposure on the auditory functions. Scand Audiol Suppl 48:63–73

Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Stat Med 9:811–818

Hong LE, Buchanan RW, Thaker GK, Shepard PD, Summerfelt A (2008) Beta (16 Hz) frequency neural oscillations mediate auditory sensory gating in humans. Psychophysiology 45:197–204

Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009) Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. Proc Natl Acad Sci U S A 106:13022–13027

Houtgast T (1972) Psychophysical evidence for lateral inhibition in hearing. J Acoust Soc Am 51:1885–1894

Houtgast T, Festen JM (2008) On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. Int J Audiol 47:287–295

Howard JH, O'Toole AJ, Parasuraman R, Bennett KB (1984) Pattern-directed attention in uncertain-frequency detection. Percept Psychophys 35:256–264

Huang MX, Edgar JC, Thoma RJ, Hanlon FM, Moses SN, Lee RR, Paulson KM, Weisend MP, Irwin JG, Bustillo JR, Adler LE, Miller GA, Canive JM (2003) Predicting EEG responses using MEG sources in superior temporal gyrus reveals source asynchrony in patients with schizophrenia. Clin Neurophysiol 114:835–850

Husain FT, Horwitz B (2006) Experimental-neuromodeling framework for understanding auditory object processing: integrating data across multiple scales. J Physiol Paris 100:133–141

Husain FT, Lozito TP, Ulloa A, Horwitz B (2005) Investigating the neural basis of the auditory continuity illusion. J Cogn Neurosci 17:1275–1292

Johnson BW, Muthukumaraswamy SD, Hautus MJ, Gaetz WC, Cheyne DO (2004) Neuromagnetic responses associated with perceptual segregation of pitch. Neurol Clin Neurophysiol 2004:33

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

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577

Kanafuka K, Nakajima Y, Remijn GB, Sasaki T, Tanaka S (2007) Subjectively divided tone components in the gap transfer illusion. Percept Psychophys 69:641–653

Kashino M, Warren RM (1996) Binaural release from temporal induction. Percept Psychophys 58:899–905

Kidd G, Mason CR, Deliwala PS, Woods WS, Colburn HS (1994) Reducing informational masking by sound segregation. J Acoust Soc Am 95:3475–3480

Kidd G, Mason CR, Rohtla TL, Deliwala PS (1998) Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. J Acoust Soc Am 104:422–431

Kidd G, Mason CR, Arbogast TL (2002) Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. J Acoust Soc Am 111:1367–1376

Kidd G, Mason CR, Richards VM (2003) Multiple bursts, multiple looks, and stream coherence in the release from informational masking. J Acoust Soc Am 114:2835–2845

Kim C, Blake R (2005) Psychophysical magic: rendering the visible 'invisible'. Trends Cogn Sci 9:381–388

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

Knight RT, Scabini D, Woods DL (1989) Prefrontal cortex gating of auditory transmission in humans. Brain Res 504:338–342

Komatsu H (2006) The neural mechanisms of perceptual filling-in. Nat Rev Neurosci 7:220–231

Kropotov J (2008) Quantitative EEG, Event-Related Potentials and Neurotherapy. Academic Press.

Larsby B, Hallgren M, Lyxell B, Arlinger S (2005) Cognitive performance and perceived effort in speech processing tasks: effects of different noise backgrounds in normal-hearing and hearing-impaired subjects. Int J Audiol 44:131–143

Larsen A, Madsen KH, Lund TE, Bundesen C (2006) Images of illusory motion in primary visual cortex. J Cogn Neurosci 18:1174–1180

Las L, Stern EA, Nelken I (2005) Representation of tone in fluctuating maskers in the ascending auditory system. J Neurosci 25:1503–1513

Leek MR, Brown ME, Dorman MF (1991) Informational masking and auditory attention. Percept Psychophys 50:205–214

Leek MR, Watson CS (1984) Learning to detect auditory pattern components. J Acoust Soc Am 76:1037–1044

Lemus L, Hernandez A, Romo R (2009) Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. Proc Natl Acad Sci U S A 106:9471–9476

Liegeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P (1994) Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol 92:204–214

Liu T, Slotnick SD, Yantis S (2004) Human MT+ mediates perceptual filling-in during apparent motion. Neuroimage 21:1772–1780

Locke RE, Nerbonne JM (1997) Role of voltage-gated K+ currents in mediating the regular-spiking phenotype of callosal-projecting rat visual cortical neurons. J Neurophysiol 78:2321–2335

Lutfi RA (1989) Informational processing of complex sound. I: Intensity discrimination. J Acoust Soc Am 86:934–944

Lutfi RA (1993) A model of auditory pattern analysis based on component-relative-entropy. J Acoust Soc Am 94:748–758

Lutfi RA, Kistler DJ, Callahan MR, Wightman FL (2003a) Psychometric functions for informational masking. J Acoust Soc Am 114:3273–3282

Lutfi RA, Kistler DJ, Oh EL, Wightman FL, Callahan MR (2003b) One factor underlies individual differences in auditory informational masking within and across age groups. Percept Psychophys 65:396–406

Lyzenga J, Carlyon RP, Moore BCJ (2005) Dynamic aspects of the continuity illusion: perception of level and of the depth, rate, and phase of modulation. Hear Res 210:30–41

Makeig S, Jung TP (1995) Changes in alertness are a principal component of variance in the EEG spectrum. Neuroreport 7:213–216

Mattys SL, Brooks J, Cooke M (2009) Recognizing speech under a processing load: Dissociating energetic from informational factors. Cogn Psychol Available at: http://dx.doi.org/10.1016/j.cogpsych.2009.04.001.

Mazaheri A, Jensen O (2008) Asymmetric amplitude modulations of brain oscillations generate slow evoked responses. J Neurosci 28:7781–7787

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

McKenna TM, Weinberger NM, Diamond DM (1989) Responses of single auditory cortical neurons to tone sequences. Brain Res 481:142–153

Meddis R, O'Mard L (1997) A unitary model of pitch perception. J Acoust Soc Am 102:1811–1820

Meng M, Ferneyhough E, Tong F (2007) Dynamics of perceptual filling-in of visual phantoms revealed by binocular rivalry. J Vis 7:8.1–815

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

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, Licklider (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

Mizuhara H, Wang L, Kobayashi K, Yamaguchi Y (2004) A long-range cortical network emerging with theta oscillation in a mental task. Neuroreport 15:1233–1238

Moore BC (1990) Co-modulation masking release: spectro-temporal pattern analysis in hearing. Br J Audiol 24:131–137

Muckli L, Kohler A, Kriegeskorte N, Singer W (2005) Primary visual cortex activity along the apparent-motion trace reflects illusory perception. PLoS Biol 3:e265

Nakajima Y, Sasaki T, Kanafuka K, Miyamoto A, Remijn G, Hoopen GT (2000) Illusory recouplings of onsets and terminations of glide tone components. Percept Psychophys 62:1413–1425

Neff DL (1995) Signal properties that reduce masking by simultaneous, random-frequency maskers. J Acoust Soc Am 98:1909–1920

Neff DL, Callaghan BP (1988) Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. J Acoust Soc Am 83:1833–1838

Neff DL, Dethlefs TM (1995) Individual differences in simultaneous masking with random-frequency, multicomponent maskers. J Acoust Soc Am 98:125–134

Neff DL, Green DM (1987) Masking produced by spectral uncertainty with multicomponent maskers. Percept Psychophys 41:409–415

Neff DL, Jesteadt W (1996) Intensity discrimination in the presence of random-frequency, multicomponent maskers and broadband noise. J Acoust Soc Am 100:2289–2298

Nielzen S, Olsson O (1997) Perceptual grouping due to pitch and amplitude in hallucinating schizophrenics. Psychopathology 30:140–148

Van Noorden 1975 Temporal coherence in the perception of tone sequences. PhD thesis

Oh EL, Lutfi RA (1998) Nonmonotonicity of informational masking. J Acoust Soc Am 104:3489–3499

Oh EL, Wightman F, Lutfi RA (2001) Children's detection of pure-tone signals with random multitone maskers. J Acoust Soc Am 109:2888–2895

Okuno HG, Ogata T, Komatani K, Nakadai K (2004) Computational Auditory Scene Analysis and Its Application to Robot Audition. ICKS '04: Proceedings of the International Conference on Informatics Research for Development of Knowledge Society Infrastructure:73–80

Oxenham AJ, Fligor BJ, Mason CR, Kidd G (2003) Informational masking and musical training. J Acoust Soc Am 114:1543–1549

Ozdamar O, Bohorquez J (2008) Suppression of the P(b) (P(1)) component of the auditory middle latency response with contralateral masking. Clin Neurophysiol 119:1870–1880

Palombi PS, Backoff PM, Caspary DM (1994) Paired tone facilitation in dorsal cochlear nucleus neurons: a short-term potentiation model testable in vivo. Hear Res 75:175–183

Patterson RD, Nimmo-Smith I, Weber DL, Milroy R (1982) The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. J Acoust Soc Am 72:1788–1803

Pekkonen E, Jääskeläinen IP, Kaakkola S, Ahveninen J (2005) Cholinergic modulation of preattentive auditory processing in aging. Neuroimage 27:387–392

Pelizzone M, Hari R, Makela JP, Huttunen J, Ahlfors S, Hamalainen M (1987) Cortical origin of middle-latency auditory evoked responses in man. Neurosci Lett 82:303–307

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

Pienkowski M, Eggermont J (2009) Effects of adaptation on spectrotemporal receptive fields in primary auditory cortex. Neuroreport Available at: http://dx.doi.org/10.1097/WNR.0b013e32832f812c.

Plack C, Oxenham A (2005) Pitch and Auditory grouping, in Pitch: Neural coding and Perception P. E. al, ed. Springer.

Pogarell O, Koch W, Schaaff N, Popperl G, Mulert C, Juckel G, Moller H, Hegerl U, Tatsch K (2008) [123I] ADAM brainstem binding correlates with the loudness dependence of auditory evoked potentials. Eur Arch Psychiatry Clin Neurosci 258:40–47

Pollack I (1975) Auditory informational masking. J. Acoust. Soc. Am Suppl. 1,57,S5

Pressnitzer Daniel RR Is the auditory continuity illusion based on a change-detection mechanism? A MEG study (unpublished, available in the proceedings of 147th Meeting of Acoustical Soceity of America, J. Acoust. Soc. Am. Volume 115, Issue 5, pp. 2460-2460, May 2004).

Pressnitzer D, Meddis R, Delahaye R, Winter IM (2001) Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus. J Neurosci 21:6377–6386

Pressnitzer D, Hupe J (2006) Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr Biol 16:1351–1357

Pressnitzer D, Sayles M, Micheyl C, Winter IM (2008) Perceptual organization of sound begins in the auditory periphery. Curr Biol 18:1124–1128

Pumplin J (1985) Low-noise noise. Acoustical Society of America Journal 78:100-104

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL (2001) A default mode of brain function. Proc Natl Acad Sci U S A 98:676–682

Rajan R, Cainer KE (2008) Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. Neuroscience 154:784–795

Ramachandran VS, Gregory RL (1991) Perceptual filling in of artificially induced scotomas in human vision. Nature 350:699–702

Razak KA, Fuzessery ZM (2008) Facilitatory mechanisms underlying selectivity for the direction and rate of frequency modulated sweeps in the auditory cortex. J Neurosci 28:9806–9816

Reese NB, Garcia-Rill E, Skinner RD (1995) The pedunculopontine nucleus–auditory input, arousal and pathophysiology. Prog Neurobiol 47:105–133

Reite M, Teale P, Zimmerman J, Davis K, Whalen J (1988) Source location of a 50 msec latency auditory evoked field component. Electroencephalogr Clin Neurophysiol 70:490–498

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

Repp BH (1992) Perceptual restoration of a "missing" speech sound: auditory induction or illusion? Percept Psychophys 51:14–32

Richards VM, Neff DL (2004) Cuing effects for informational masking. J Acoust Soc Am 115:289–300

Riecke et al Hearing illusory sounds in noise: the timing of sensory-perceptual transformations (unpublished, available in the procedeengs of at the Auditory cortex 2009 conference).

Riecke L, Opstal AJV, Goebel R, Formisano E (2007) Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. J Neurosci 27:12684–12689

Riecke L, Orstal AJV, Formisano E (2008) The auditory continuity illusion: a parametric investigation and filter model. Percept Psychophys 70:1–12

Roberts B, Glasberg BR, Moore BCJ (2008) Effects of the build-up and resetting of auditory stream segregation on temporal discrimination. J Exp Psychol Hum Percept Perform 34:992–1006

Rosburg T, Trautner P, Korzyukov OA, Boutros NN, Schaller C, Elger CE, Kurthen M (2004) Short-term habituation of the intracranially recorded auditory evoked potentials P50 and N100. Neurosci Lett 372:245–249

Rose MM, Moore BC (1997) Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. J Acoust Soc Am 102:1768–1778

Rose MM, Moore BC (2000) Effects of frequency and level on auditory stream segregation. J Acoust Soc Am 108:1209–1214

Ross B, Herdman AT, Pantev C (2005) Stimulus induced desynchronization of human auditory 40-Hz steady-state responses. J Neurophysiol 94:4082–4093

Samuel AG (1981) Phonemic restoration: insights from a new methodology. J Exp Psychol Gen 110:474–494

Sanchez-Morla EM, Garc?a-Jimenez MA, Barabash A, Martinez-Vizcaino V, Mena J, Cabranes-Diaz JA, Baca-Baldomero E, Santos JL (2008) P50 sensory gating deficit is a common marker of vulnerability to bipolar disorder and schizophrenia. Acta Psychiatr Scand 117:313–318

Sarter M, Hasselmo ME, Bruno JP, Givens B (2005) Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. Brain Res Brain Res Rev 48:98–111

Schwindt PC, Spain WJ, Crill WE (1989) Long-lasting reduction of excitability by a sodium-dependent potassium current in cat neocortical neurons. J Neurophysiol 61:233–244

Seeba F, Klump GM (2009) Stimulus familiarity affects perceptual restoration in the European starling (Sturnus vulgaris). PLoS One 4:e5974

Shahin AJ, Bishop CW, Miller LM (2009) Neural mechanisms for illusory filling-in of degraded speech. Neuroimage 44:1133–1143

Singh PG, Bregman AS (1997) The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. J Acoust Soc Am 102:1943–1952

Snyder JS, Alain C (2007a) Sequential auditory scene analysis is preserved in normal aging adults. Cereb Cortex 17:501–512

Snyder JS, Alain C (2007b) Toward a neurophysiological theory of auditory stream segregation. Psychol Bull 133:780–799

Stapells DR, Linden D, Suffield JB, Hamel G, Picton TW (1984) Human auditory steady state potentials. Ear Hear 5:105–113

Stopfer M, Laurent G (1999) Short-term memory in olfactory network dynamics. Nature 402:664–668

Sugita Y (1997) Neuronal correlates of auditory induction in the cat cortex. Neuroreport 8:1155–1159

Sumbre G, Muto A, Baier H, Poo M (2008) Entrained rhythmic activities of neuronal ensembles as perceptual memory of time interval. Nature 456:102–106

Thurlow W (1957) An auditory figure-ground effect. Am J Psychol 70:653–654

Tougas Y, Bregman AS (1990) Auditory streaming and the continuity illusion. Percept Psychophys 47:121–126

Trainor LJ, Trehub SE (1989) Aging and auditory temporal sequencing: ordering the elements of repeating tone patterns. Percept Psychophys 45:417–426

Ulanovsky N, Las L, Farkas D, Nelken I (2004) Multiple time scales of adaptation in auditory cortex neurons. J Neurosci 24:10440–10453

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398

Warren RM (1970) Perceptual restoration of missing speech sounds. Science 167:392–393

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, Hainsworth KR, Brubaker BS, Bashford JA, Healy EW (1997) Spectral restoration of speech: intelligibility is increased by inserting noise in spectral gaps. Percept Psychophys 59:275–283

Warren RM, Obusek CJ, Ackroff JM (1972) Auditory induction: perceptual synthesis of absent sounds. Science 176:1149–1151

Warren R (1999) Auditory perception: A new Analysis and Synthesis, 2nd edition. Cambridge University Press.

Warren R (2008) Auditory perception: A new Analysis and Synthesis, 3rd edition. Cambridge University Press.

Watson CS, Kelly WJ, Wroton HW (1976) Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. J Acoust Soc Am 60:1176–1186

Weerd PD, Smith E, Greenberg P (2006) Effects of selective attention on perceptual filling-in. J Cogn Neurosci 18:335–347

Weerd PD (2006) Perceptual filling-in: More than the eye can see. Prog Brain Res 154:227–245

Wehr M, Zador AM (2005) Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47:437–445

Weisser R, Weisbrod M, Roehrig M, Rupp A, Schroeder J, Scherg M (2001) Is frontal lobe involved in the generation of auditory evoked P50? Neuroreport 12:3303–3307

Wightman FL, Kistler DJ (2005) Informational masking of speech in children: effects of ipsilateral and contralateral distracters. J Acoust Soc Am 118:3164–3176

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

Winer JA (2006) Decoding the auditory corticofugal systems. Hear Res 212:1–8.

Winter Ian M, Neuwert, V, Verhey JL (2005) Comodulation masking release and the role of wideband inhibition in the cochlear nucleus, in the Auditory Signal Processing: Physiology, Psychoacoustics, and Models. D. Pressnitzer, ed. Springer Verlag.

Yan W, Suga N (1998) Corticofugal modulation of the midbrain frequency map in the bat auditory system. Nat Neurosci 1:54–58

Yarrow K, Haggard P, Heal R, Brown P, Rothwell JC (2001) Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. Nature 414:302–305

Zekveld AA, Kramer SE, Vlaming MSMG, Houtgast T (2008) Audiovisual perception of speech in noise and masked written text. Ear Hear 29:99–111

Zhang X, Carney LH (2005) Analysis of models for the synapse between the inner hair cell and the auditory nerve. J Acoust Soc Am 118:1540–1553

Zokoll MA, Klump GM, Langemann U (2008) Auditory memory for temporal characteristics of sound. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 194:457–467

Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. Annu Rev Physiol 64:355–405

Supplementary material

Appendix 1. Supplementary figures

Chapter 5

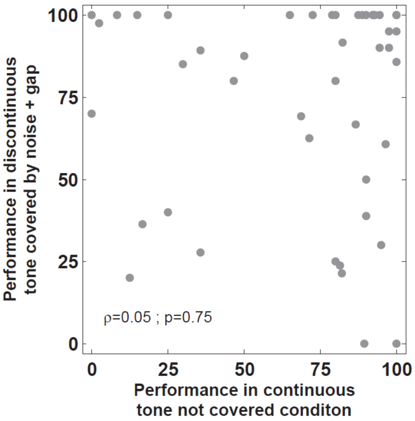


Figure S5.1 A. Relation between performance in the continuous tone not covered by noise condition (data from 50 ms noise durations) and performance in the discontinuous tone covered by noise + gap condition (data from 2000 ms noise durations, N=46) This effect appears to be independent from subjective disruption of continuity because participants that perform above chance in the continuous tone conditions do not differ from the others, (p= 0.2, rank sum statistic: 375, N=46 (18 and 28 in each group)).

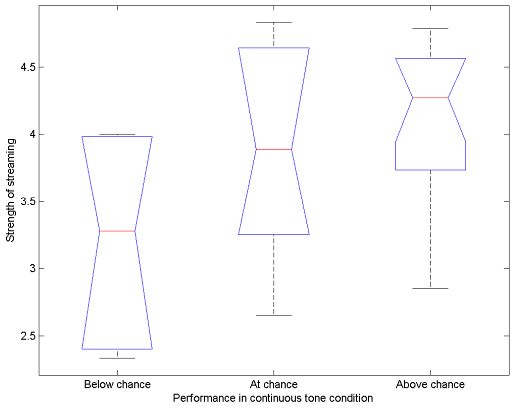


Figure S5.2 Streaming data grouped according to 3 levels of performance in CTNC (below chance, at chance, above chance).

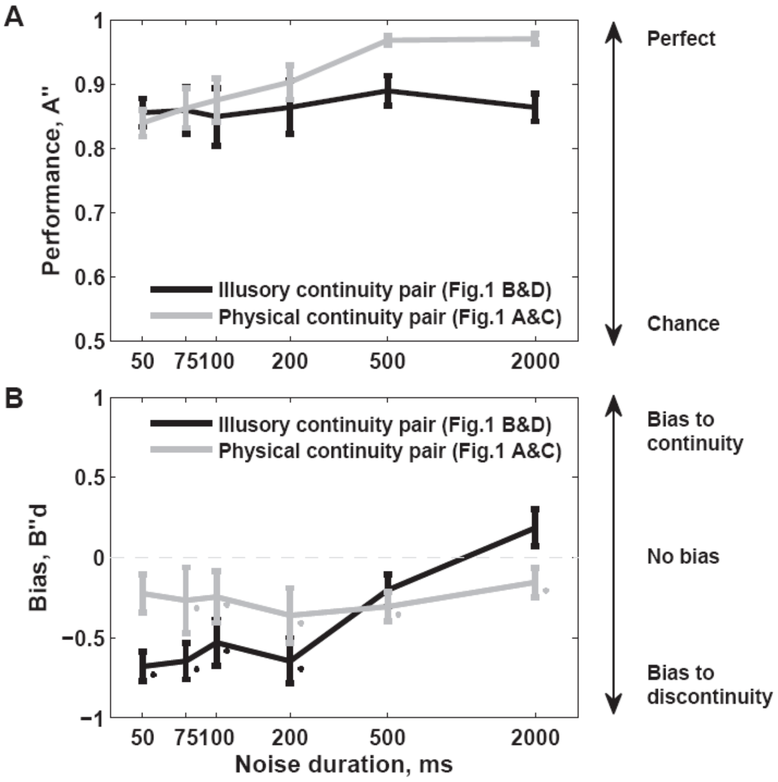


Figure S5.3 A. Bias-free sensitivity (A’) in illusory and physically continuous pairs of conditions, plotted against noise duration on the X axis. **B. Bias** (B’’d) in participants’ responses. Asterisks indicate presence of significant bias in the responses (p<0.001, Bonferroni-corrected) at particular noise durations. The bias to discontinuity is stronger in the case of illusory pair conditions compared to physically continuous pair conditions at short noise durations. (N=46 for noise durations of 50, 500, and 2000 ms, N=18 for 75, 100 and 200 ms, mean ± SE)

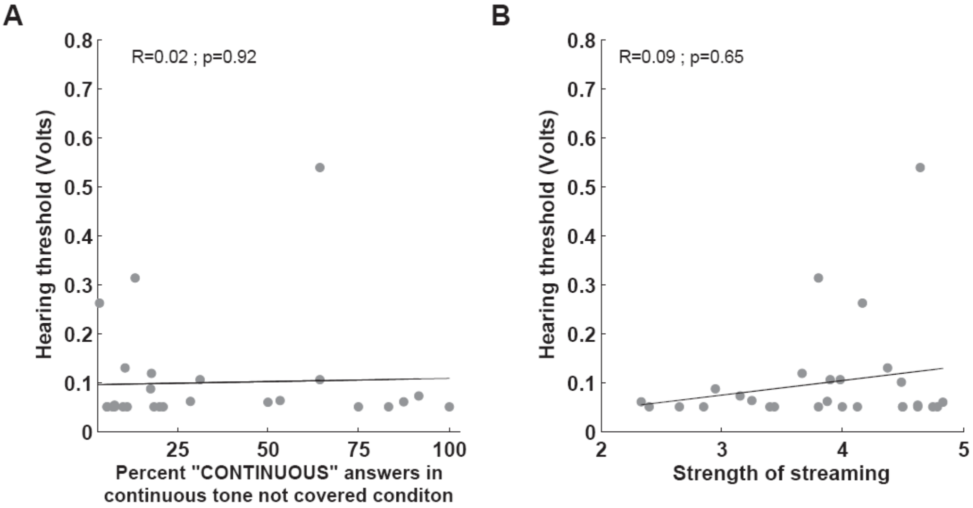


Figure S5.4. Tone detection in noise thresholds and subjective disruption of continuity (no significant correlation, Spearman test, N=18). **B** Hearing thresholds and results of the streaming test (no significant correlation, Spearman test, N=18).

Chapter 6

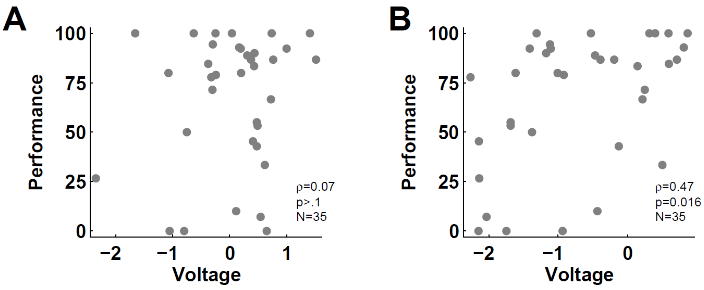


Figure S6.1 Individual susceptibility to discontinuity illusion and ERP to a 50-ms noise presented alone in two time windows: **A.** 40-66 ms; **B.** 270-350 ms. Each dot reflects the measures from a single individual. Spearman’s ρ and p values (corrected for multiple comparisons) are indicated.

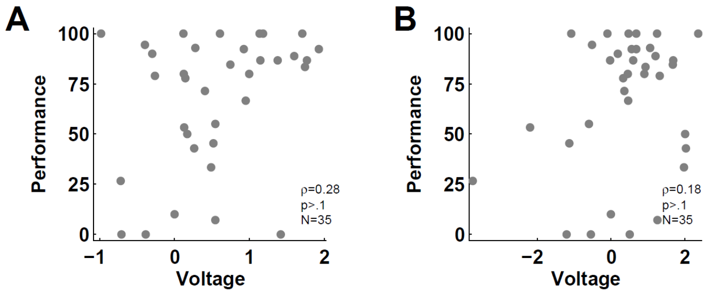


Figure S6.2 Individual susceptibility to discontinuity illusion and ERP to a tone plus 1000-ms noise presented alone in two time windows: **A.** 40-66 ms; **B.** 270-350 ms. Each dot reflects the measures from a single individual. Spearman's ρ and p values (corrected for multiple comparisons) are indicated.

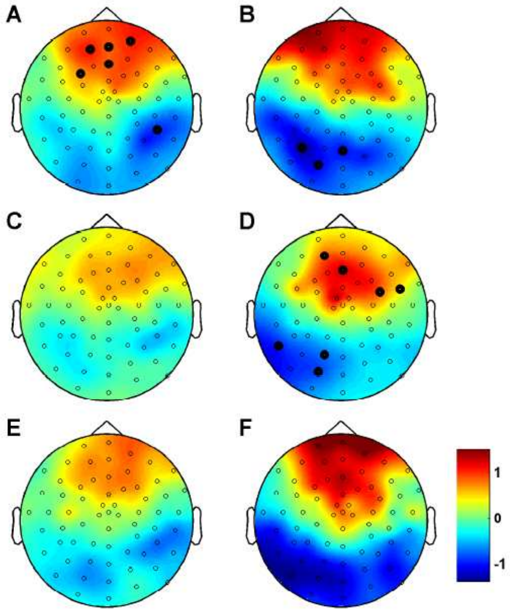
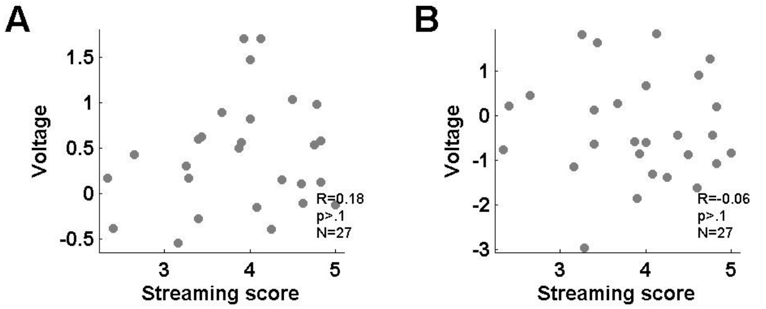


Figure S6.3 Scalp distribution of the potential differences between participants that exhibit a bias toward continuity (N=20) vs. those that exhibit a bias toward discontinuity (N=6). **Left column** – 40-66 ms integration window; **right column** – 270-250 ms window. **A, B** – 50 ms noise and tone condition; **C, D** – 50 ms noise alone; **E, F** – 1000 ms noise and tone. Red color signifies electrodes with bigger values the "continuity" group compared to the "discontinuity" group. Electrodes with correlation between performance and voltage (p<0.005, uncorrected) are highlighted.

Figure S6.4 Individual streaming scores in a sequential task (Experiment 2) and ERP voltage in the integration windows: A. 40-66 ms after noise onset; **B** 270-350 ms after noise onset. Each dot reflects the measures from a single individual. Pearson correlation coefficient and p values (not corrected for multiple comparisons) are indicated.



Chapter 7

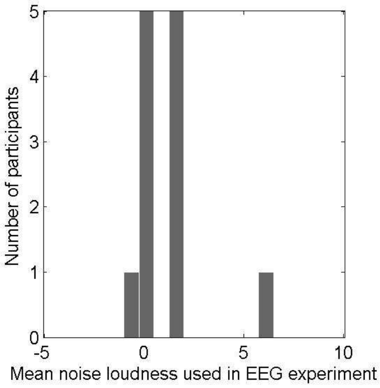


Figure S7.1 Individual noise loudness used in experiment 4 in dB relative to tone level. N=12, for 2 of the participants there was no data available.

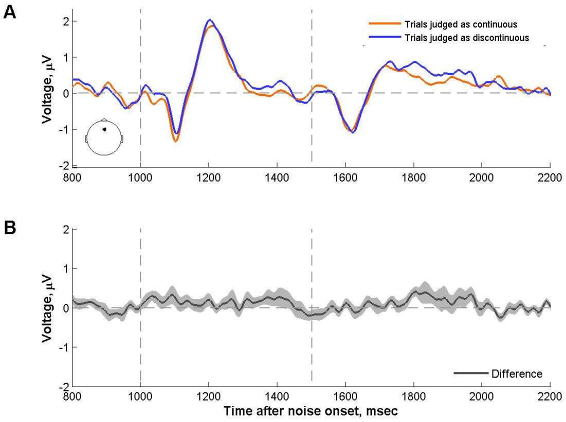


Figure S7.2 Evoked potentials in the trials judged as continuous and discontinuous, aligned to the onset of the noise. **A.** Grand averages (N=14); the dashed lines denote the times of noise onset and offset. **B.** Difference waveform. The line and surrounding shaded area reflect the mean and the standard error of the participants' difference waveforms (trials judged as discontinuous minus trials judged as continuous).

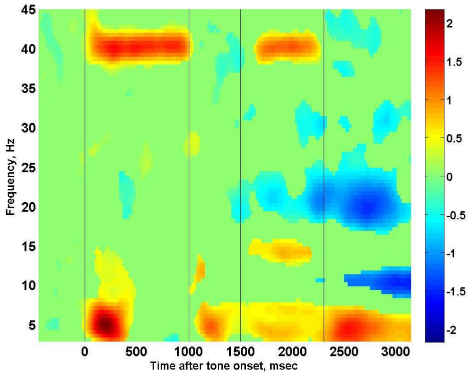


Figure S7.3 Event-related spectral perturbations (ERSP) at the Fz electrode (all trials pooled together irrespective of perceptual judgement). The black lines denote (from left to right) the onset of the tone, the onset of the noise, the offset of the noise, the offset of the tone. There is no tone present during the noise. The ERSP is thresholded by the nominal p-value according to paired t-tests (comparing each point with a baseline, p<.05, N=14).

Appendix 2. A method for generating sounds with a notch in the modulation spectrum

This method may be used to modify natural sounds or synthesize new sounds with desired modulation spectra. It was used in the pilot studies on EEG correlates of the continuity illusion (data not reported in the thesis). The idea was to create a masking noise which specifically lacks amplitude modulation at 40 Hz and thus will not interact with the neuronal coding of 40 Hz amplitude modulation of the tone.

The initial approach was to minimize amplitude modulation of the noise, using the original method of Pumplin (1985), based on a gradient search procedure to find a set of phases for all components of the sound that would result in minimal amplitude modulation, or, alternatively, to use a simpler method to produce “low-noise noise” (used in Dau et al, 1999). It is an iterative technique: on each iteration the noise is band-pass filtered and divided by its own envelope. The resulting signal has very low modulation levels. However, these methods are suited only for very narrow-band sounds. For the wider-band cases (e.g. the 600 Hz wide band-passed noise used here) the procedure described above produces large distortions of spectrum and timbre so that the noise does not sound heterogeneous, but rather has bursts of lower frequencies interleaved with bursts of higher frequencies. Jon Pumplin, Emily Buus and Sergio Callegari are thanked for generously providing clarifications and sharing their scripts for generating “low-noise noise”.

The procedure was modified to eliminate not all the modulation frequencies from the noise but only a narrow band surrounding the target modulation frequency of 40 Hz. On each iteration, the signal was divided by its envelope filtered from 30 Hz to 50 Hz with a linear phase delay filter.

The script for the algorithm is provided here:

```
% make sample of noise

          LOW=randn(1,npts);

          tmp_F=fft(LOW).*Mask;

          LOW=real(ifft(tmp_F));

 % make Noise with a notch in a modulation spectrum (LOW)

          for loop=1:3,

                    % filter the envelope

                              filtered_envelope=filter_30_50_hz_zero_phase_delay(abs(hilbert(LOW)),FS)
                    +mean(abs(hilbert(LOW)));

                    % divide the signal by the filtered envelope

                               LOW=LOW./filtered_envelope;

                    % filter the signal with the band-pass filter

                              tmp_F=fft(LOW).*Mask;

                              LOW=real(ifft(tmp_F));

          end
```

The resulting sound has homogenous timbre very close to the original band-passed noise; however, some blurring of the spectrum is present.

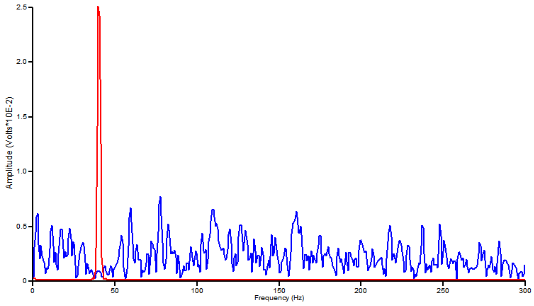


Figure S8.1 Modulation spectrum of a noise with a notch around 40 Hz (blue) and a modulation spectrum of a tone, amplitude modulated at 40 Hz.