ABSTRACT

Title of Document: AUDITORY EDGE DETECTION:
THE DYNAMICS OF THE CONSTRUCTION
OF AUDITORY PERCEPTUAL
REPRESENTATIONS

Maria Chait, Ph.D., 2006

Directed By: Prof. D. Poeppel, Departments of Linguistics and
Biology
and
Prof. JZ Simon, Departments of Biology and
Electrical and Computer Engineering

This dissertation investigates aspects of auditory scene analysis such as the detection

of a new object in the environment. Specifically I try to learn about these processes

by studying the temporal dynamics of magnetic signals recorded from outside the

scalp of human listeners, and comparing these dynamics with psychophysical

measures. In total nine behavioral and Magneto-encephalography (MEG) brain-

imaging experiments are reported. These studies relate to the extraction of tonal

targets from background noise and the detection of change within ongoing sounds.

The MEG deflections we observe between 50-200 ms post transition reflect the first

stages of perceptual organization. I interpret the temporal dynamics of these

responses in terms of activation of cortical systems that participate in the detection of

acoustic events and the discrimination of targets from backgrounds. The data shed

light on the statistical heuristics with which our brains sample, represent, and detect

changes in the world, including changes that are not the immediate focus of attention. In particular, the asymmetry of responses to transitions between 'order' and 'disorder' within a stimulus can be interpreted in terms of different requirements for temporal integration. The similarity of these transition-responses with commonly observed onset M50 and M100 auditory-evoked fields allows us to suggest a hypothesis as to their underlying functional significance, which so far has remained unclear. The comparison of MEG and psychophysics demonstrates a striking dissociation between higher level mechanisms related to conscious detection and the lower-level, pre-attentive cortical mechanisms that sub-serve the early organization of auditory information. The implications of these data for the processes that underlie the creation of perceptual representations are discussed. A comparison of the behavior of normal and dyslexic subjects in a tone-in-noise detection task revealed a general difficulty in extracting tonal objects from background noise, manifested by a globally delayed detection speed, associated with dyslexia. This finding may enable us to tease apart the physiological and behavioral corollaries of these early, pre-attentive processes. In conclusion, the sum of these results suggests that the combination of behavioral and MEG investigative tools can provide new insights into the processes by which perceptual representations emerge from sensory input.

.

AUDITORY EDGE DETECTION:
THE DYNAMICS OF THE CONSTRUCTION OF AUDITORY PERCEPTUAL
REPRESENTATIONS

By

Maria Chait

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Ph.D.
2006

Advisory Committee:
      Professor David Poeppel   Co-chair
      Professor Jonathan Z. Simon Co-chair
      Professor Colin Phillips
      Professor Catherine E Carr
      Dr. Jonathan Fritz
      Professor Robert J Dooling

The brain is wider than the sky,
  For, put them side by side,
The one the other will include
  With ease, and you beside.

The brain is deeper than the sea,
  For, hold them, blue to blue,
The one the other will absorb,
  As sponges, buckets do.

The brain is just the weight of God,
  For, lift them, pound for pound,
And they will differ, if they do,
  As syllable from sound.

Emily Dickinson (1830-86)

I dedicate this dissertation with all my love
to my grandmother Riva Hodun and my mother Valentina Chait.

Acknowledgements

Table of Contents

Introduction

The many sound-generating sources in the environment produce an aggregate waveform that enters the ear. In order to operate in, and make sense of the world, a listener has to be able to separate this input into source-related components, localize them, recognize them, and react appropriately. These processes mostly occur automatically and without explicit effort: we are able to hold a conversation despite the TV being on, follow the voice of a singer amidst the background music or accurately estimate which instruments are playing in a band.

In this dissertation I will discuss what can be learned about this auditory perceptual organization by investigating how brain responses to sounds unfold over time. During the course of auditory processing, sensory information undergoes a transformation from a representation in purely physical terms to an abstract, perceptually relevant form - two identical signals may be apprehended completely differently depending on the state of the perceiver. The way in which a listener perceives auditory input depends on task and listening mode (what s/he is listening for), previous experience (e.g. 'sine wave speech'; Remez et al., 1994) and may even change randomly during listening (bi-stability phenomena; e.g. Gutschalk et al., 2005; Pressnitzer & Hupé, 2005).

Since perception is inherently private and automatic, it is difficult to tap into its constructing stages. In my work, I employ simultaneous psychophysical and brain-imaging tools (magnetoencephalography; MEG) to study the processes that underlie a listener's construction of a representation of the acoustic environment. With its fine

temporal resolution, MEG is especially useful for investigating the temporal dynamics of how these representations arise from sensory input and task demands. In the following chapters I will discuss the mechanisms by which auditory objects are detected and separated from background and the role of top-down processing such as attention.

In chapter 1 I detail the advantages and limitations of the MEG technique. One important constraint is that we are only able to resolve cortical activity. Before reaching this stage, sensory information has already undergone a great deal of processing. Some would argue that most features of sound have already been identified, analyzed and represented by the time activity arrives in auditory cortex. Nevertheless studying auditory cortical processing is extremely interesting because it is believed (e.g. Nelken 2004) to be the stage at which sound is no longer represented in terms of its physical properties but is converted into behaviorally relevant representations in terms of 'auditory objects'. With MEG we can track the gross changes in brain activation as stimuli appear or change, compare certain aspects of how they are represented and also study the effects of attention.

In this dissertation I am focusing on very early, pre-attentive brain responses occurring between 50-200ms post onset. I show, however, that bottom up effects already have complex influences on cortical brain responses. That is, how we listen to sounds affects in profound ways even relatively early, pre-conscious, stages of processing.

In the experiments described here I, use very simple stimuli: pure tones or wide band noise, which lack the richness of natural sounds, yet share certain basic

characteristics with those signals. The hope is that through analyzing brain activation in response to these impoverished signals, we may acquire insight as to how more complex signals are processed, and which dimensions of sound might be perceptually relevant.

The first stage in analyzing a visual scene is the detection of edges. It seems that in audition too the first stages of analyzing a complex auditory scene include detecting edges between sound sources. This is the topic of this dissertation. Edges are complex concepts. They don't just define the boundaries of objects but in fact may define how objects are represented neurally and how they are perceived. The experiments in the following chapters examine the temporal dynamics of the cortical systems that subserve the detection of auditory temporal edges - the neural basis of listener's ability to make sense of an ever changing, complex acoustic world.

Chapter 1: What are the data?

'The tragedy of the fisherman from the North Sea', as heard in a lecture by Prof. Shlomo Biderman in the course 'Introduction to religion and philosophy in India', Tel Aviv University, Fall 1998)

> *This is a story of an old fisherman who lived by the North Sea. His family has fished the icy waters for generations. He had never traveled anywhere, the sound of the waves and the smell of the salt in the air were his constants. He didn't really care much about other places on the globe, his comfort being that **this place** – the North Sea he knows well. He was an expert seaman, able to read the sea and the sky, and their respective creatures like books that tell him what he needs to know. Then once, somebody asked him a question he couldn't answer: "what is the size of the smallest fish in the North Sea?" At first he was astounded, the question never occurred to him before. The realization that he does not have a response was uncomfortable and disconcerting – and he made it the rest of his life's mission to find out the answer. He consulted with old books, tried the fishery organization and then created a plan. It took years to execute. He meticulously fished in various areas of the sea, diverse depths, numerous times of day and different seasons. He cautiously sorted the catch in his net, measured the fishes and carefully noted the sizes in his notebook, doing everything as precisely as he could. His neighbors thought he was crazy. He was growing older, pushing the boat into the sea was becoming harder every morning, hauling the nets up was a laborious and demanding process, starting before dawn and often lasting until midday. But his mission is what warmed him in freezing winds and lashing rain. It took 10 years to complete all the measurements he set out to perform. And at the end of the last trip, when the last haul was inspected and noted, he knew he had the answer. He leafed through his notebook looking for the smallest number, and there is was- 1.5 cm. The size of the smallest fish in the North Sea. His undertaking was over. The feeling of calmness that he remembered from 10 years ago – the confidence of being the master of his surroundings was back. He went to fold his net. But his joy soon turned into a terrible tragedy – it was when he was holding his net, he noticed that the diameter of the squares in the mesh was exactly 1.4 centimeters.*

This dissertation concerns MEG measurement and the dynamics of brain activation, and what they can teach us about the processes by which perceptual representations

are created. I devote this first chapter to a discussion of what the recorded MEG data represent. This is important because the limits of the data constrain the questions we can ask, and the answers we can obtain from this experimental methodology.

1.1 What is the MEG signal?

1.1.1 Raw data

At the end of the 19th century, when the electrical activity of the brain was beginning to be discovered, a major issue was to establish that the signals were indeed originating from the central nervous system and are not motor or tissue artifacts (Tcheriev, 1902; Tcheriev, 1904; Brazier, 1961). Today, the origin of signals measured in EEG and MEG is better understood, but there remain many missing links, particularly in relating the macroscopic electric or magnetic fields to the microscopic activity (single or multiple units) commonly observed in animal physiology. This is partly because the latencies we record (on the time scale of 100ms post stimulus onset) are so much beyond the latencies commonly observed in animal physiology. Current work in animal EEG/MEG, as well as pre-surgically implanted electrodes in humans (e.g. Liegeois-Chauvel et al, 2004; Bares et al, 2003; Godey et al, 2001; Gueguin et al, 2006) is instrumental in bridging the gap between these levels. Nevertheless, there is still some uncertainty as to what neural process the MEG signals represent.

MEG measures magnetic fields outside the head produced by electrical activity in neural cell assemblies within the brain. Measurable magnetic fields are

thought to originate from excitatory post synaptic currents (EPSCs) in the apical dendrites of tens of thousands of simultaneously active cells (Hari et al., 2000; Baillet et al., 2001; Kaufman & Lu, 2003). On the assumption that the source of the magnetic field is a single segment of current that is small relative to the distance at which the sensors are located, the source can be modeled as an "equivalent current dipole" (ECD). It is commonly stated that pyramidal cells are the major source of the MEG signal because of their size and parallel arrangement (Nunez & Silberstein, 2000; Baillet et al., 2001) however this assumption is under debate (e.g. Kaufman & Lu, 2003).

As compared to other techniques to measure brain activity (such as PET or fMRI), MEG, by virtue of being a direct measure of electrical activity in the brain, has fine temporal resolution (on the order of ms; restricted only by digitization rate), but limited (~5mm-1cm) spatial resolution. This is a consequence of the relatively small number of sensors (spatial channels), and the inherently ambiguous nature of the electromagnetic inverse problem. Given the magnetic field distribution outside the head, it is impossible to compute where the underlying sources are located, because different source configurations may produce the same field. Useful inferences on the location of sources are possible only by applying restrictive source models. As compared to EEG, MEG has several advantages: (1) MEG offers better spatial resolution because the skull and brain tissue are 'transparent' to magnetic fields and MEG thus avoids the smearing effect of the high-impedance skull and multiple conductive tissue layers on head surface potential distributions. Without any restrictive models MEG allows gross separation of activity into hemispheres, as well

as the analysis of finer topographical details, and more effective source modeling. (2) The auditory cortex, located in the Sylvian fissure of the temporal lobe, is oriented so that currents flow mainly parallel to the head surface. This puts it in the "blind spot" of EEG, but also, as it were, in the "fovea" of MEG. MEG is particularly sensitive to currents tangential to the skull. (3) MEG is easier set up and less unpleasant for the subject (no need for conductive cream to enhance electrode-skin contacts). The disadvantage of MEG over EEG is its greater cost, both in terms of investment (costly sensors and shielding), and maintenance (need for a supply of liquid helium, calibration, etc.).

1.1.2 Analysis

The research described in this dissertation is based on the properties of "stimulus locked" responses, averaged over several hundreds of repetitions of the same auditory stimulus. The raw MEG signals, produced by the (in our case, 157) SQUID sensors distributed over the head, are segmented into 'epochs' (including the stimulus and short pre-stimulus and post-stimulus intervals) and averaged over repetitions. The averaging operation reduces environmental noise and background brain activity, and keeps brain activity components that are repeatable and precisely time-locked to the stimulus ('evoked response'). It is important to realize that this procedure removes potential activity that might be elicited by the stimulus, but with a time-course that is not precisely locked to it ('induced response').

The earliest evoked magnetic responses peak at about 20, 30, and 50 ms post stimulus onset (conventionally designated respectively as M20, M30 and M50). Collectively referred to as "middle latency responses" (Woods et al., 1995; Yvert et

al., 2001; Rupp et al., 2002; Lütkenhoner et al., 2003), these responses are thought to originate in or close to primary auditory cortex (Yvert et al, 2001). They are followed by a peak at about 100ms (M100) whose current source has been localized to the upper banks of the superior temporal gyrus (Hari, 1990; Pantev et al, 1995; Lütkenhöner & Steinsträter, 1998)[1]. Because of its prominence and robustness across subjects and stimuli, the M100 has been the most investigated brain response in relation to auditory processing (Naatanen & Picton, 1987; Roberts et al., 2000). Figure 1.1 illustrates a typical MEG response to a steady state pure tone (tone onset is at 0 ms) that is 1400ms long. Each line represents the activity recorded by one sensor. The typical M50, M100 and M150 (as well as later responses) are evident as peaks in activity. Figure 1.2 shows the MEG response recorded in response to one of our stimuli described in Chapter 3, Section 2. The stimulus consisted of a steady-state pure tone that changed into a sequence of random frequency tone pips (transition at 840 ms post-onset). In addition to the onset response described above, we see a response to the onset of the random tone sequence. What kind of processing do these deflections reflect?

1.2 Interpretation of the data

As discussed above, auditory evoked responses result from synchronous, stimulus time-locked activity in tens of thousands of cortical cells (Hämäläinen et al., 1993). What could generate such coordinated increases in current? A pessimistic hypothesis

---

[1] Some models suggest that the M100 has at least 2-3 underlying sources (e.g. Picton et al., 1999; Scherg & Von Carmon, 1985)
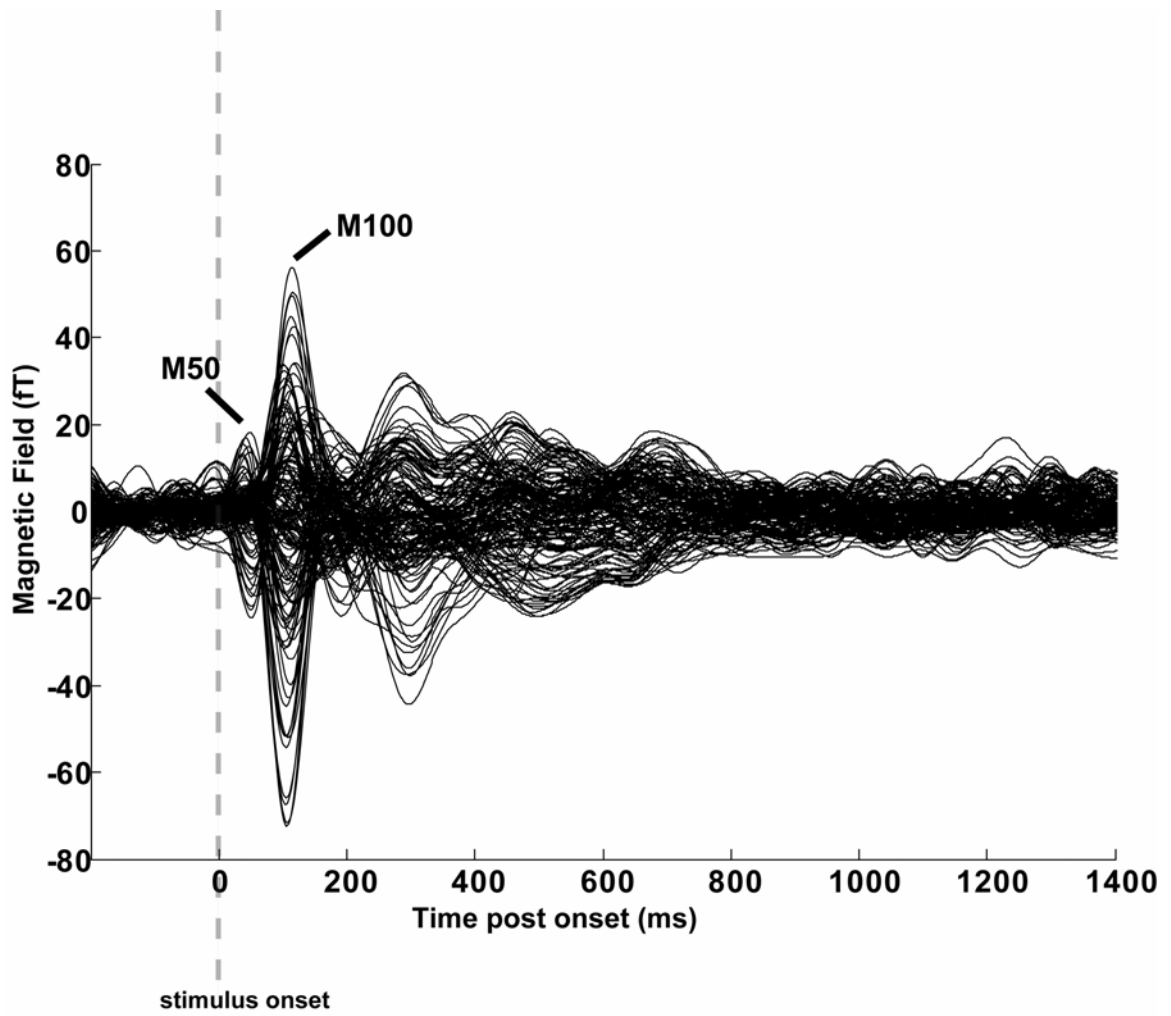
Figure 1.1: MEG response to a 1400ms long pure tone (onset at 0 ms). Each line represents the magnetic field measured by one sensor. Highlighted are classic onset responses M50 and M100.
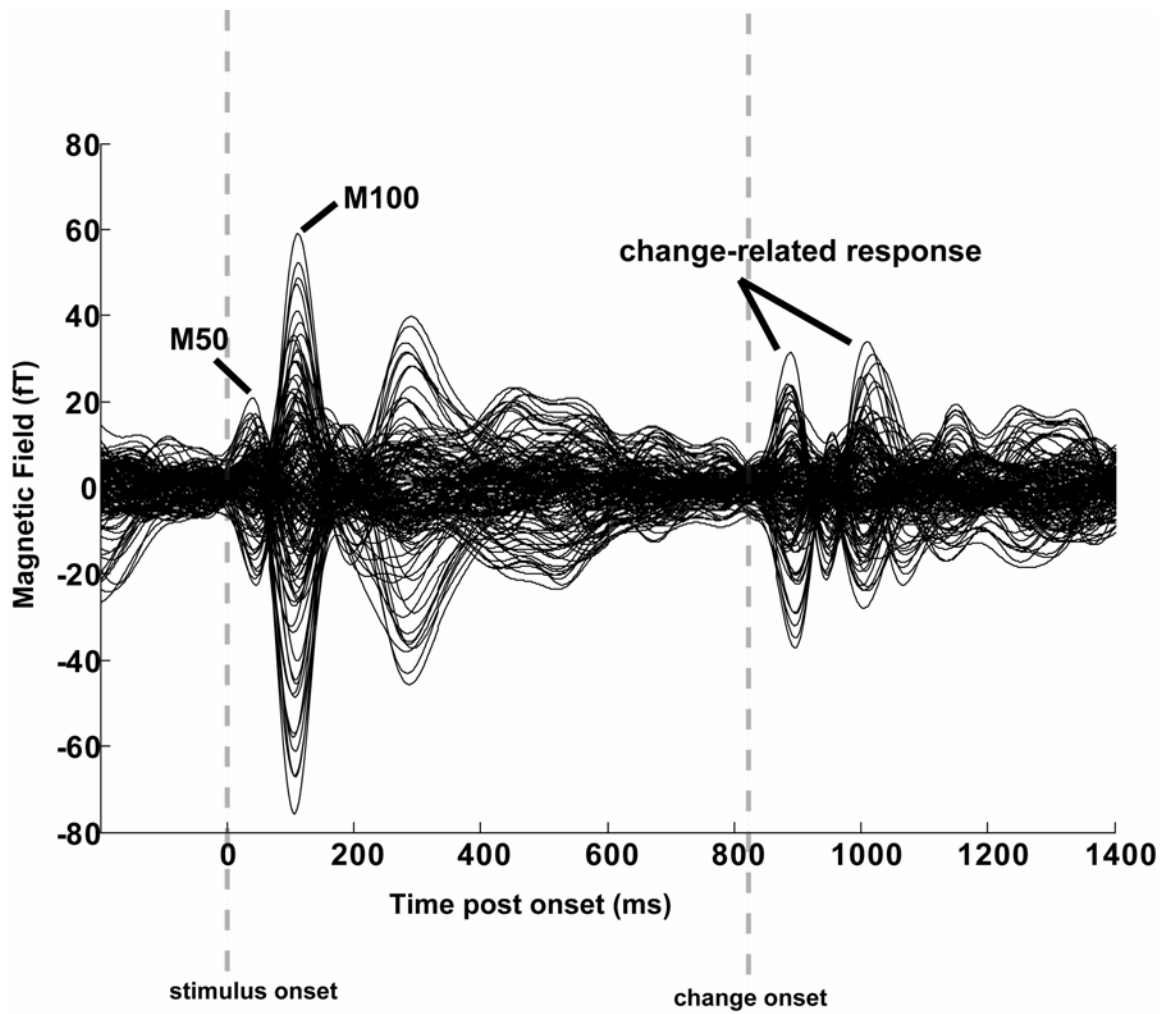
Figure 1.2: MEG response to a pure tone which changes (at 840 ms) to a sequence of 60ms tone pips of random frequency.

is that they do not reflect actual information processing, but rather operations related to general resource management, akin for example to *garbage collection* within a software system (e.g. Abelson et al., 1996). Pushing further this metaphor, suppose that we were to put a computer, rather than a brain, into the MEG machine. Activity related to information processing would likely remain invisible because of the small currents involved and the complexity of patterns. Instead, larger events such as disk-head movements would most certainly show up, on a background of ongoing magnetic activity due to the spinning disk, dynamic RAM refreshing, switching power supply, etc. We might however discover that this 'disk-head response' is reliably triggered by events that cause a momentary peak in information processing activity (as a result of memory swaps or disk buffer pages being marked as 'dirty'). We could then claim that the magnetic responses measured from outside the computer are telling us something about information processing within it. We are, potentially, in the same situation with respect to the brain. Indeed, one can argue that onset responses such as M50 or M100 occur later than expected from computations associated with processing stimulus attributes and that similar responses are found in other sensory modalities, with very different stimulus processing demands. The time at which memory needs resetting, and the amount of resetting required, might depend on the timing and nature of the stimulus, and thus the response might seem to reflect details of the mechanisms that process the sensory information. This "garbage-collection" hypothesis is perhaps over-pessimistic, but it is worth keeping it in mind when we discuss details of MEG stimulus-response characteristics.

Another interpretation, referred to as the phase resetting model or event-related-synchronization (Penny et al., 2002; Klimesch et al, 2004; Gruber et al, 2005), suggests that onset responses do not reflect an increase in neural activity per se but are generated by a stimulus evoked phase-resetting of ongoing cortical oscillations. However, this issue is currently under debate (e.g. Mazaheri & Jensen, 2006).

The most common interpretation of event-related brain responses, and the one that will be adopted here as a working hypothesis, is that this activation pattern, consisting of increases in activation at different latencies, reflects the sequential operation of cortical systems related to analyzing the stimulus (e.g. Friederici, 2005).

What kinds of mechanisms are likely to be 'visible' in the auditory evoked signal? Recent studies have interpreted auditory evoked components as reflecting the output of a pitch processing mechanism (e.g. Krumbholz et al., 2003; Ritter et al., 2005; Seither-Preisler et al., 2006). Others have interpreted them as reflecting timbre, location, etc. (Soeta et al., 2005b; Mäkelä et al, 2003; see Roberts et al, 2000 for review). In each case, the interpretation is based on the observed relation between response characteristics (such as peak latency) and stimulus parameters (such as period). However at least two arguments can be made against such interpretations. One is that, if all these interpretations are correct, the "code" (e.g. peak latency) is awfully overloaded. In other words, the response feature might code one or the other among these features, but not all. The second argument, related to my analysis of the computer metaphor, is that it is unlikely that processing of a feature such as pitch would require the simultaneous and synchronized activation of tens of thousands of cells. *A priori*, the computations that I consider to be more likely to evoke the

observed MEG responses are more global mechanisms related to object analysis, notification of change, attentional switching etc. (Chait et al., 2004; Gutschalk et al., 2004b). Conceivably, such processes may involve the synchronous activation of many cells as a method of notification across mechanisms and brain areas that something new and potentially behaviorally relevant has occurred in the environment. This will be the main theme of the work described here.

1.3 Phenomenology of event related brain responses

Most MEG experiments, and indeed all of those reported here, involve comparing cortical responses to stimuli that differ along some dimension. What can be learned from such a comparison? A first outcome is obviously that a stimulus dimension that affects a response dimension is "relevant" in that it has a demonstrable effect on brain state or activity. However, we usually want to go further than this, and attribute the difference to a *representation* of the stimulus. That is, we would like to infer that the magnetic deflection that differed between the two stimuli somehow reflects the processing of the relevant stimulus feature. It is not debatable that changes in this activation pattern (as a result of modifying the stimulus or the listener's attentional state) indicate changes within the neural mechanisms that the pattern reflects. Variations in latency may be interpreted as a slowing down or speeding up of a particular process, likewise changes in amplitude may reflect a difference in processing demands. The difficulty is in deciding to what aspect of the stimulus, or the state of the listener, these patterns should be attributed.

A change in amplitude is particularly hard to interpret. A greater amplitude in condition A vs. condition B may be interpreted as meaning that the processing in A is easier (Pantev et al, 2003; Bosnyak et al, 2004; Shahin et al, 2003). However it is just as plausible to attribute it to an increase in processing demands (Kutas & Feithermeir, 2000; Pylkkanen et al, 2004; Friederici, 2005). The appropriate interpretation depends on the processing model one has in mind. *Latency* differences are perhaps more straightforward to interpret as reflecting the order of processing stages, or the amount of time that a given processing stage requires. For that reason, latencies will be the focus of discussion in this dissertation.

A further issue is the level of description at which MEG responses are interpreted. 'Cognitive' studies, concerned with higher order functions such as memory or language, tend to interpret a deflection as reflecting a (theory dependent) macroscopic mechanism underlying a certain psychological function. Studies concerned with lower-level sensory information tend to interpret responses in more microscopic terms of neural activation. As discussed above, the data we record are quite remote from the single cell level, and thus such microscopic interpretations are dangerous. Discussion of MEG response properties in terms of 'receptive fields', 'refractory periods', 'firing rates' or 'temporal coding' may be misleading because those terms apply to a different level of description. The effects which we observe on the evoked response level may be implemented neurally in various different ways (e.g. Budd et al., 1998; Grill-Spector et al, 2006). In this argument, we are basically repeating the point made by Marr (1982) that the different levels of description on

which brain phenomena are discussed or interpreted must be carefully kept distinct.

1.4 Effects of stimulus presentation

The ultimate goal in auditory experimentation is to understand how the brain processes sound in *natural* conditions. The lab setting is obviously different from a listener's everyday environment. Methodological constraints on the manner in which stimuli are presented can have severe effects on the resulting brain activity, and on the conclusions one draws from the data. The experiments that we describe reveal multiple examples of such (not altogether expected) dependencies. Two specific issues are stimulus *repetition* and *randomization*.

Averaging responses over multiple repetitions of the same stimulus is a common technique for increasing the signal to noise ratio (SNR) of the data. All the experiments to be presented in the following chapters involve repetitions of essentially the same stimulus for about an hour. It is known that stimulus repetition has strong effects on brain activity: response amplitudes typically become weaker, and such attenuation may affect some components more than others, thus resulting in a distorted response signature (e.g Roth et al., 1976; Budd et al, 1998; Naatanen & Picton, 1987; Grill-Spector, 2006). In experiments involving different stimuli, this can lead to order effects that masquerade as stimulus dependencies. To reduce the likelihood of order-related artifacts, stimuli are usually presented in random order, or the order of presentation is counterbalanced across subjects (as in chapter 3 here). However it is important to keep in mind that the activity we might be tapping on to is
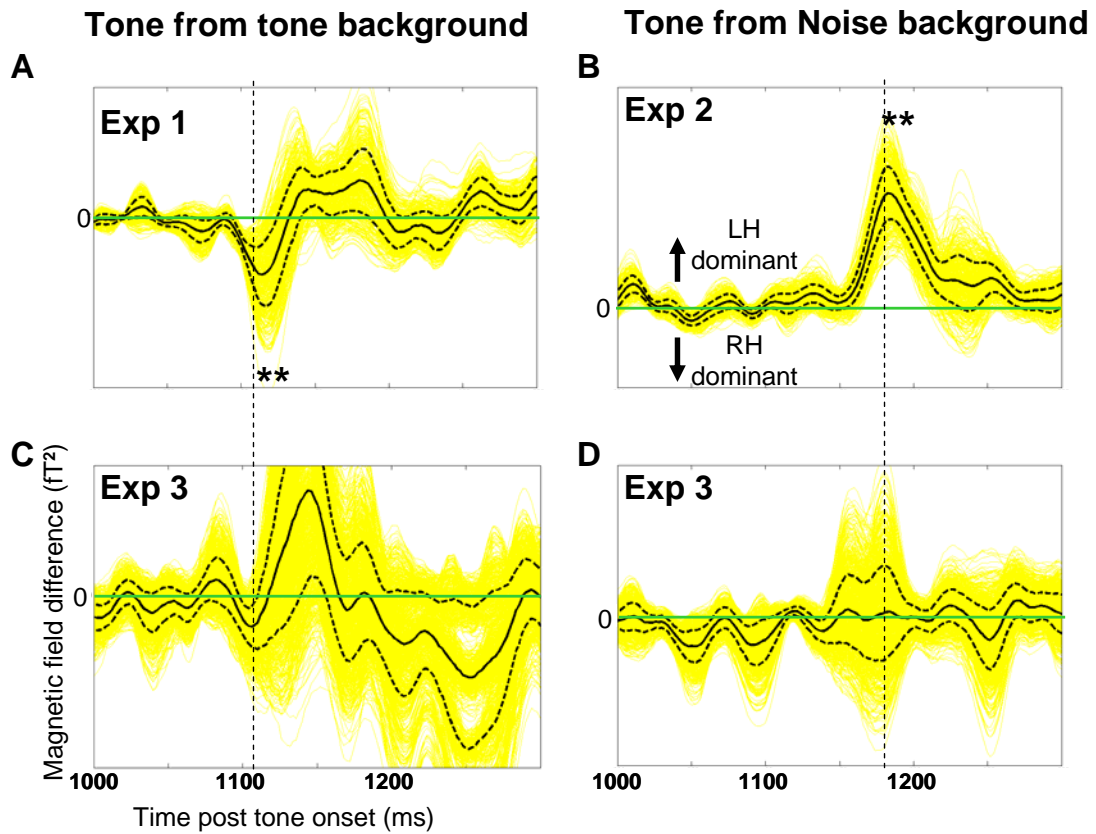
Figure 1.3: Laterality of MEG responses (difference of power between left and right hemisphere) to the onset (at 1000 ms) of a tone within a tone (A,C) or noise (B,D) background. The continuous black line is the mean over subjects, the yellow lines are bootstrap trials, and the dotted lines indicate one standard deviation of the bootstraps from the mean. A and C: responses to identical tone-on-tone stimuli in different context, in Exp1 (blocked) and Exp 3 (interleaved). B and D: responses to identical tone-on-noise stimuli in Exp2 (blocked) and Exp3 (interleaved). Vertical dotted lines are the latency of the tone-onset peak in the group RMS (RMS of individual RMSs) data. The large differences (top vs. bottom) show the importance of the stimulus context in which responses are gathered.

to a large extent influenced by the fact that stimuli are repeated and stimulus repetition per se may have a very big effect on the mechanisms we end up observing (e.g. Dyson et al., 2005).

In addition to repetition, the wider context within which the stimuli are presented may have strong effects on what we measure. In a recent study we compared data from three separate experiments that used the exact same stimuli, presented in different contexts (Figure 1.3). The experiments were designed to investigate the neural processes which underlie the extraction of a tonal object from a background. In Exp 1 of that study, the background was a pure tone. In Exp 2 it was white noise. Exp 3 was a replication of Exp 1 and Exp 2, in which listeners heard stimuli from both experiments interleaved in a randomized fashion. All the stimuli that appeared in Exp 3 also appeared in one or the other of the first two experiments. The task in all three experiments was identical: "respond when you detect a tone popping out of the background". Comparing responses to the same stimuli between Exp 1 and Exp 3 (left column of Fig 1.3), and between Exp 2 and Exp 3 (right column) reveals major differences: Whereas a significant right lateralization was found in Exp 1, it is absent in Exp 3. Whereas a strong left lateralization was found in Exp 2, it is also absent in Exp 3. Statistical analyses (bootstrap) indicate that the effect is unlikely to result from differences between subjects used in Exp 1 and Exp 2 on one hand, and Exp 3 on the other. We conclude that these differences are likely to be the result of the different stimulus context in which stimuli were presented (blocked in Exp 1 and 2, interleaved in Exp 3).

1.5 Effects of task and attention

Yet another issue is the effect of the subject's attentional state. In the experiments to be described here, subjects are required to perform a task while in the MEG machine. Depending on the experiment, the task may play several very different roles. In some cases (as in Chapter 2), the task involves the stimulus feature for which brain responses are being measured. The purpose is to measure how MEG activation is related to behavioral performance. In other cases (as in Chapter 3), the task instead involves some stimulus feature that is irrelevant to that for which the brain responses are gathered. The purpose in that case is to keep the subjects awake and attentive to the auditory modality, but without engaging the stimulus features of interest for the study by either attention or behavior. Typically, responses are elicited for a small proportion of stimuli (e.g. 30%). Brain responses to those stimuli are usually discarded. In the first case, the focus is on the attentive processing that leads to a behavioral response. In the second case, the focus is more on pre-attentive and bottom-up aspects of processing. By "pre-attentive", we mean processes that monitor ongoing acoustic information, whether or not that information is involved in conscious perception or behavior. This notion is discussed in more detail in Chapter 3.

Our distinction between attentive and pre-attentive processes does not imply that the latter are not affected by attentional state or task. Early cortical auditory processing is not an impenetrable module (e.g Fodor, 1983), and there is accumulating evidence that the task performed by a subject can exert a very strong effect on activity at all levels of the auditory system, down to the cochlea. It is well

known that brain activation for the same stimulus differs according to whether subjects are listening passively or required to process the information to perform a task (Zatorre et al, 1992; Poeppel et al., 1996; Jancke et al., 1999; Brechmann & Scheich, 2005). It is less well known that the nature of the task affects processing of unattendend stimuli. Although there is a rich literature on this in the visual domain.

One outcome of this thesis research is precisely that the effects of task are more widespread and complicated than usually reported. For example, most of the stimuli used in the experiments presented here are relatively long (~1500 ms) and consist of an initial segment, which subjects usually are not required to attend to, and a change segment on which a task may be focused. In Chapter one, these stimuli are noise segments with a tone appearing at the end, In section 1 of Chapter 3 these are noise segments which become modulated towards the end. We discovered that responses to the onset of these sounds differ significantly from responses to onset of the same physical sounds under passive listening conditions. This conclusion differs from previous reports on effects of attention, because here in both cases subjects were equally attentive to the stimuli, and in both cases no specific response was required to the onset of stimuli. This is discussed further in Chait et al (2004). Figure 1.4 here provides an example of such responses. The stimuli are 990Hz pure tones. The experiment, designed to probe task effects, consisted of three stages (see figure 2.22 in Chapter 2 here for a description of the stimuli): (1: PRE) a pre-experiment where subjects listened to approximately 200 repetitions of a 1000 ms 990 Hz pure tones and asked to count the signals, (2:MAIN) Main experiment: Stimuli consisted of 1500 ms 990 Hz pure tone with a 500 ms faint tonal object appearing at 1000 ms post
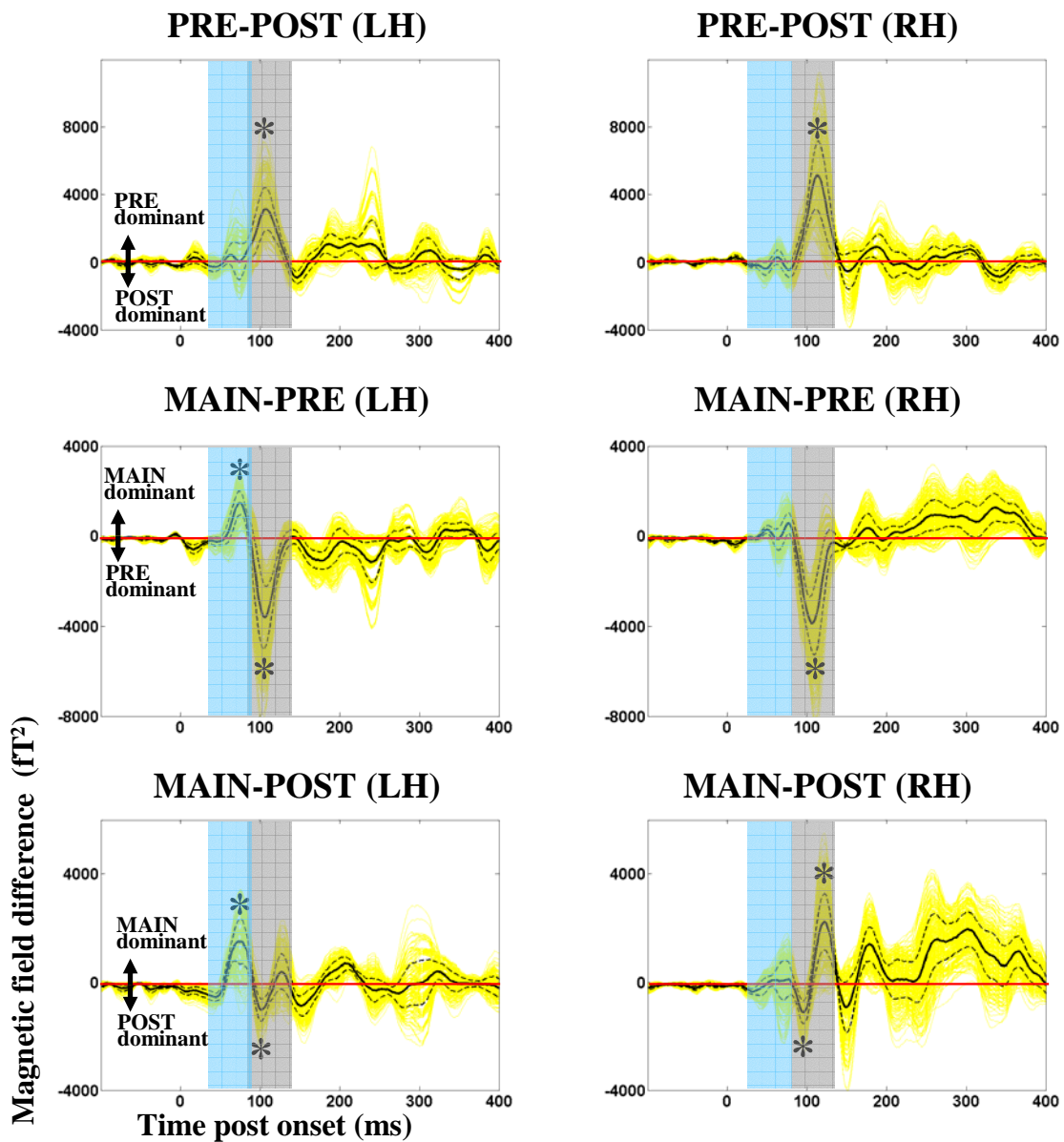
Figure 1.4: Comparison of onset responses to PRE, POST and MAIN conditions in the left and right hemispheres. Plotted are differences in activation between pairs of conditions The continuous black line is the mean over subjects, the yellow lines are bootstrap trials, and the dotted lines indicate one standard deviation of the bootstraps from the mean. M50 and M100 time windows are marked with blue and grey shadings, respectively. Responses to PRE and POST do not differ except that the M100 amplitude is significantly attenuated in POST relative to PRE. This is the classic effect of habituation of ERP responses. However, the comparisons with the MAIN condition reveal a complex pattern of differences. This differences which is in consistent with explanations in terms of habituation is hypothesized to reflect the effect of task demands on early cortical activation.

onset. The control stimulus was a 1500 ms 990 Hz pure tone. (3: POST) At the end of the experiment subjects again listened to approximately 200 repetitions of 1000 ms long 990 Hz tones, and were asked to count the signals. Figure 1.4 compares onset responses to each of these conditions. Plotted are the *differences* in activation between pairs of conditions, for each hemisphere. The initial part of the stimulus was physically identical in all three conditions, and one might expect the responses to be the same (the difference plots should be flat). Instead, we see large systematic differences. Comparison of PRE and POST reveals stronger activation for the former in the M100 time window in both hemispheres. Such habituation of the M100 during an experiment is a robust finding in the literature.

However, comparing the responses of MAIN to either PRE or to POST reveals a complex pattern of differences that can not be explained by habituation: During the MAIN section of the experiment there is an increase in activation in the M50 time window (blue shading): the M50 peak is of higher amplitude during MAIN than in the proceeding PRE or the following POST sections. Additionally, there is a decrease in the M100 activation in the MAIN Section relative to PRE, which, as stated above may be exclusively due to habituation, however, if habituation is the only cause we would expect M100 activation to continue decreasing during the POST stage. Conversely what we see, in fact, is a moderate *increase* in M100 time window in POST relative to PRE (especially in the left hemisphere). The increase in M100 amplitudes in the POST conditions and M50 amplitudes in the MAIN condition is a pattern that contradicts what would be expected from habituation. I believe that the relevant factor behind this difference in response patterns is the fact that, whereas in

21

PRE and POST the subjects had no task to perform, in MAIN they performed a task that engaged their attention to a later part of the stimulus, after the onset.

Brain activity as early as putative primary auditory cortex (50ms post sound onset) is already modulated by the status assigned to the sound by the listener. This demonstration differs from other auditory attention experiments (such as Woldorff et al., 1993) in that these employ spatial selective attention: subjects are instructed to listen to signals in one ear, while ignoring the opposite ear. In our paradigm subjects are not specifically instructed to ignore any aspect of the scene. Instead the focus of attention is defined, more abstractly, by the requirements of the task. When subjects perform a task that is related to the later part of the stimulus, the former part is relegated to the background (or perhaps: resources are diverted from the stimulus onset). When subjects listen passively, pre-attentive mechanisms treat the stimulus onset as a potential target and assign resources by default. These conclusions are consistent with effects of auditory attention measured from human electrocorticograms (Neelon et al., 2006) as well as reports from the visual domain where cell activation beginning at about 40ms and peaking at about 100ms is consistently correlated with the status of the input as figure or ground (Lamme, 1995; Zipser et al., 1996; Super et al, 2001; Heinen et al, 2005).

Another way in which a subject's task may affect responses is depicted in Figure 1.5. Figure 1.5 plots the activation for the interaural-correlation stimuli used in Chapter 3. The aim of the experiment was to measure brain responses to changes in interaural correlation of a stimulus that was not attended to. To keep them awake, subjects performed a task that required them to detect the onset of amplitude
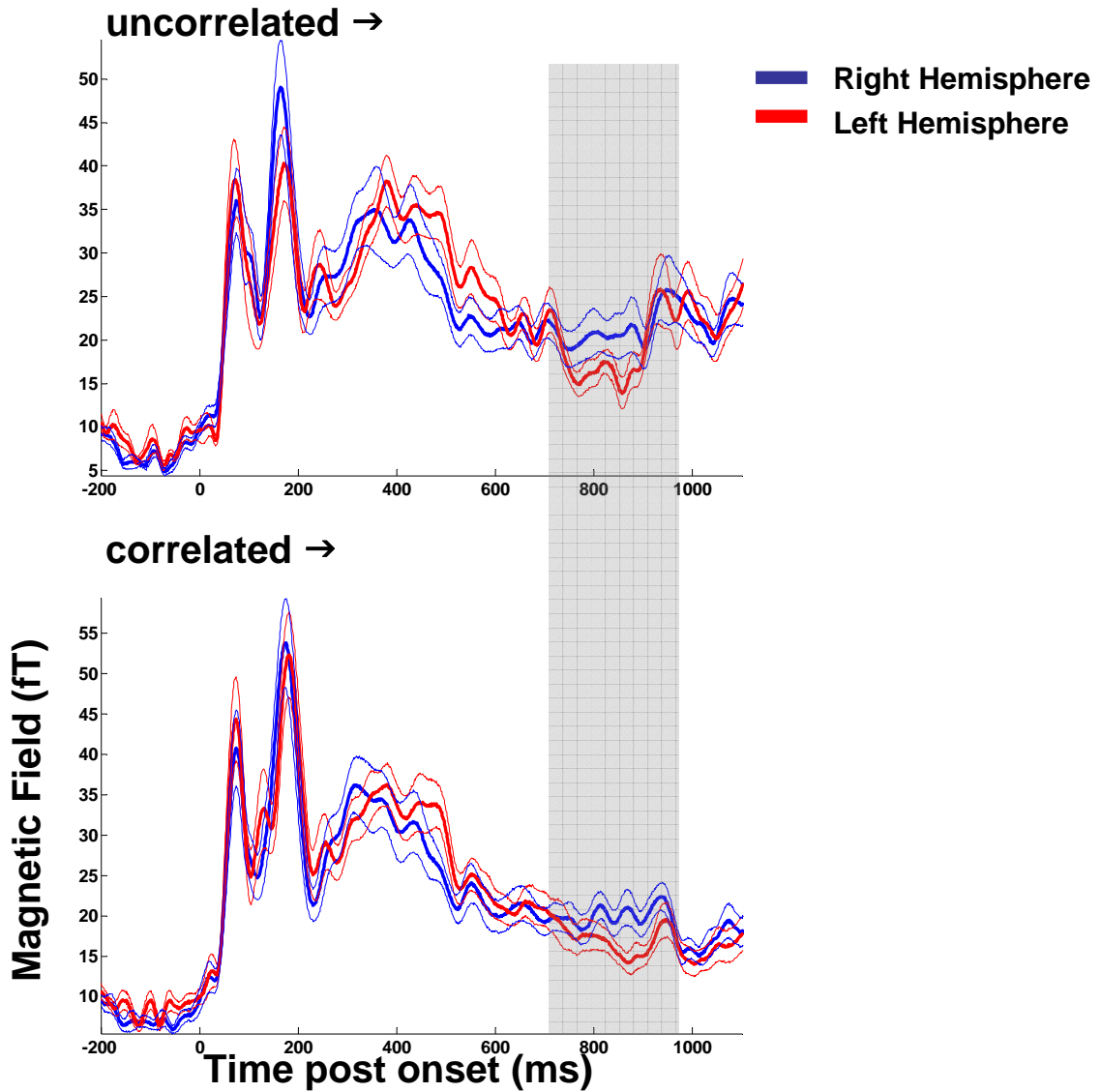
Figure 1.5: Activation time course for the interaural correlation stimuli in Chapter 3. Plotted are data for the two compound conditions (see methods section Chapter 3.1). In both hemisphere there is a strong response to the onset of the stimulus, that progressively dies down towards 800 ms. However, the right hemisphere shows a less strong decline than the left hemisphere, and there appears to be a plateau starting slightly before 800 ms which reflects anticipation of amplitude modulation onset.

modulation of ongoing noise at 800ms post onset. Brain responses for these behavioral target stimuli were discarded from analysis. The task did *not* require processing of interaural correlation. However it appears nevertheless that the task induced a listening mode that did affect our results. Figure 1.5 compares activation time courses in left (red) and right (blue) hemispheres for the non-target (interaural correlation change) stimuli. In both hemispheres there is a strong response to the onset of the stimulus, that progressively dies down towards 800 ms. However, the right hemisphere shows a less strong decline than the left hemisphere, and there appears to be a plateau starting slightly before 800 ms which reflects anticipation of the change in the stimulus. Right hemisphere advantage for processing AM in this range of modulation rate has been documented (Zatorre, 1992; Zatorre 2000). This in turn might be the factor behind the right hemisphere dominance that we observed for processing changes in interaural correlation (Chait et al, 2005; Chapter 2 here). Such effects of listening mode are a possible confound in comparisons between conditions for which listening mode is not rigorously controlled. They might also lead to conclusions about processing (e.g. 'the right hemisphere has an advantage in processing changes in interrural correlation') that are in fact unjustified.

All of this is to say that we need to be careful that what we are measuring is indeed what we think we are measuring, and aware of the possibility that it is not. The data in the following chapters need to be evaluated and interpreted under the limitations and caveats described here.

1.6 Conclusions

MEG gives the researcher an opportunity to glimpse inside the brain of a human listener and follow, potentially with millisecond resolution, the activation triggered by the stimulus. However the relation between the observed response and the underlying processes is unclear, and may be affected in unexpected ways by seemingly benign experimental factors. Part of the difficulty comes from that MEG responses, despite their many channels and high temporal resolution, are too crude to reflect the richness of activity within the brain. As with any other technique, it is important to be aware of these limitations and ask experimental questions that are answerable in a satisfactory way so as not to spend our time measuring our net instead of the size of the smallest fish in the North Sea.

Chapter 2: Processes underlying the extraction of tonal objects from background noise

Our everyday environment is noisy, consisting of many concurrent sound sources that appear and disappear at random. In order to make sense of the world a listener has to be sensitive to the emergence of new, possibly relevant, auditory objects, be able to localize them, recognize them and react accordingly. A first step in this process is the detection of events within the ongoing background noise.

In this chapter I report a series of MEG and psychophysical experiments that measure how listeners detect the appearance of tonal targets amidst background noise. We compared the cortical and behavioral responses to the onset of Huggins Pitch (HP), a stimulus requiring binaural processing to elicit a pitch percept, with responses to tones embedded in noise (TN)—perceptually similar but physically very different signals. The comparison of these two kinds of targets is attractive because it allows to tease apart peripheral vs. central contributions to the cortical representation of the signal: Aspects of the responses to the two stimuli that are qualitatively identical would imply central processing (probably above the Inferior Colliculus), whereas differences can be attributed to peripheral processing (see also Julesz, 1971). Such an evaluation is useful because it puts constraints on processing models and the locus of relevant computations.

This chapter is comprised of three main sections. Section 1, which provides the experimental data for the other two sections, is based on Chait et al. (2006) and

appears here in its published form. Subjects listened to stimuli consisting of tonal objects that appeared occasionally, embedded in wideband noise, and were instructed to press a button as fast as they can when they heard a tone. Although this was an active detection task, the responses we observed were essentially identical to other reports with similar stimuli where recordings were made under passive listening conditions (Krumbholz et al, 2003; Seither-Preisler et al., 2004).

The paper is worded in terms of pitch processing because it appears in the context of the 'Pitch Onset Response' (POR) literature (Krumbholz et al., 2003). However it is likely that the responses we observed reflect a more general process related to the detection and extraction of an object from the background noise (in this case giving rise to a pitch). There are several arguments in favor of this conclusion: The hemispheric lateralization of the tone onset-response is inconsistent with that observed in pitch related tasks (Zatorre, 2001; see also Hall et al, 2005). The properties of the response are similar those of the 'object related negativity' reported by Alain and colleagues (Alain et al., 2002). A third line of evidence will be developed in Chapter 3.

There is a fourth argument that I mention here briefly but do not discuss further in this thesis. In analyzing the data from these experiments, we discovered an unusual pattern of response to the onset of our stimuli, different from that usually reported for such signals, and indeed observed by us in the context of other tasks (Chait et al, 2004; Chait et al., 2005). Specifically, we observed a change in the early onset response (discussed in Chapter 1): an increased amplitude for M50 and decreased amplitude for M100. This change occurred when the task required subjects

to focus their attention on a later portion of the stimulus, suggesting that the initial noise may have been processed as a background. We have performed follow-up experiments that target this effect, but the results are still being analyzed and will not be reported here.

Sections 2 and 3 are investigating how these early, pre-attentive, tone-onset responses are related to perception. In Section 2, we look more closely at certain properties of responses observed in the above experiments. By comparing the time course of the behavioral responses and the cortical responses in the MEG signal, we are able to construct a model of the processing stages involved in extracting the tonal targets and relate pre-attentive target segregation with conscious detection.

The third section, describes a study with dyslexic listeners using the same stimuli. Previous investigations, employing stimuli similar to HP, reported that dichotic pitch detection is significantly impaired in individuals with developmental dyslexia as compared to average readers (Dougherty et al., 1998). These results have been interpreted to suggest a low-level auditory deficit associated with Dyslexia. Our HP and TN stimuli and the behavioral paradigm introduced in Chait et al (2006) are a useful tool to address the issue of dyslexics' alleged impaired binaural processing, and to differentiate between relatively high level impairments (for example, difficulties in extracting objects from noise) from low level deficits, related to combination of fine temporal information across ears. An important result of Chait et al (2006) was that (normal) subjects responded (behaviorally and electrophysiologically) approximately 30ms faster to HP stimuli than to TN stimuli, regardless of frequency tested. This latency difference is attributable to low level

binaural processes (as early as the medial superior olive) that operate differently on the two kinds of signals. A low level deficit associated with dyslexic listeners should therefore result in an altered pattern of latency differences between HP and TN stimuli, as compared to normal reading controls.

The data, analyzed in this way, show no evidence of a specific binaural impairment in dyslexia. However, dyslexic subjects exhibited a general difficulty in extracting tonal objects from background noise, manifested by a globally delayed detection speed. In section 3, I discuss how further study of this population may turn to be an intriguing opportunity to investigate the mechanisms responsible for detection of targets in noise and the patterns of behavioral and electrophysiological deficit that result from the malfunction of these neural systems. In particular, it may shed light on how these early cortical processes are related to perception.

2.1 Neural response correlates of detection of monaurally and binaurally-created pitches[2]

Pitch, one of the most salient features evoked by sound, is crucial to our ability to process voiced speech, segregate auditory streams, and enjoy music. Despite the importance of pitch, the mechanisms responsible for its extraction, as well as their location in the brain, are a matter of debate. The information necessary for the computation of pitch of some signals is available as early as the auditory nerve of either ear (Moore, 1997; Cariani & Delgutte, 1996). Others require combination of information from both ears within a central pitch processor. Houtsma and Goldstein (1971) demonstrated that the 'missing fundamental' effect (a harmonic complex tone has a pitch determined by its repetition rate even if a sinusoidal component at that frequency – the fundamental - is not physically present in the signal) can occur even when harmonics are played to different ears. This suggests that the 'pitch extractor' for such stimuli resides at or above the level of the Superior Olivary Complex (SOC), where the information from the two ears is first combined. On the other hand, binaural interaction is not required for other stimuli and it has been suggested that both monaural and binaural pitch mechanisms might exist (e.g. Carlyon et al., 2001; see reviews in Moore, 1997; de Cheveigné, 2005).

Pitch processing in humans has primarily been studied via psychophysics (see Moore, 1997; Plack et al., 2005). These studies have been augmented by brain

---

[2] This section is a reproduction of Chait M, Poeppel D, Simon JZ. (epub 2005) Neural Response Correlates of Detection of Monaurally and Binaurally Created Pitches in Humans. Cereb Cortex.

imaging research carried out with fMRI/PET (Griffiths et al., 1998a; Patterson et al., 2002; Penagos et al., 2004) and EEG/MEG (Pantev et al., 1996a; Fujioka et al., 2003; Krumbholz et al., 2003; Gutschalk et al., 2004; Ritter et al., 2005). Evidence from several studies (Patterson et al., 2002; Penagos et al., 2004; Ritter et al., 2005) points to an area immediately antero-lateral to primary auditory cortex (PAC) as the area where pitch extraction-related processes may operate. These have recently been complemented by similar findings in animal electrophysiology (Bendor & Wang, 2005).

Whereas fMRI and PET are well suited to address questions related to *where* processing occurs, MEG excels in the investigation of the *time course* of processing. In the present study we combine a crafted auditory stimulus, *Huggins Pitch* (HP), with the high temporal resolution of MEG recording, to investigate aspects of the processing of pitch information in cortex. Specifically, we investigate the *timing* of the response to tonal targets within background noise. Huggins Pitch (Cramer & Huggins, 1958) is the auditory counterpart of the more famous 'random dot stereogram' (RDS) effect in vision (Julesz, 1971). An RDS is a binocularly presented pair of images with randomly distributed dots. Each image appears featureless when viewed individually but together they evoke a coherent 3-D percept when displayed simultaneously, one to each eye. The illusion is created by presenting identical RDSs to the two eyes except that one image contains a group of dots that are slightly shifted relative to the other. The visual system fuses the shifted and non-shifted dots to create a 3D percept of an image (corresponding to the shifted dots) floating above the background (of the un-shifted dots). Similarly in audition, if a random broadband

noise signal is presented to one ear, and the same random noise—but with a phase shift of π over a narrow frequency band—is presented to the other ear, this results in the perception of a faint tonal object with a pure tone quality (with a pitch that matches the center frequency of the phase-shifted band), embedded in noise (Fig. 2.1). It is crucial that the input to either ear alone is just white noise, completely lacking any spectral or temporal cues to pitch. The fact that we are nevertheless able to perceive pitch when the two signals are presented dichotically implies that the HP percept is created by a central mechanism that receives the inputs from the two ears, collates them, and derives from the correspondence the percept of a tone. Here we compare the cortical auditory evoked responses to HP to a stimulus that is physically different but nonetheless elicits a very similar percept: a pure tone embedded in noise (TN).

For our tonal stimuli, we chose four frequencies, ranging from 200 to 1000 Hz. The HP signals were generated by inverting a narrow spectral region of a noise sample in one ear, centered about the tonal frequency, while the original sample is delivered to the other. The corresponding TN signals were produced by adding a pure tone (with amplitude chosen to match the perceived tone loudness of the corresponding HP stimulus) to the original noise signal. Matched TN and HP stimuli result in a very similar perceptual experience.

Experimentally it is important to isolate the processing that is specific to the detection of the onset of a tonal stimulus from that associated with generic stimulus onset. Typically this involves comparing responses to noise and tone-like stimuli. However the large onset component common to both reduces the sensitivity to the
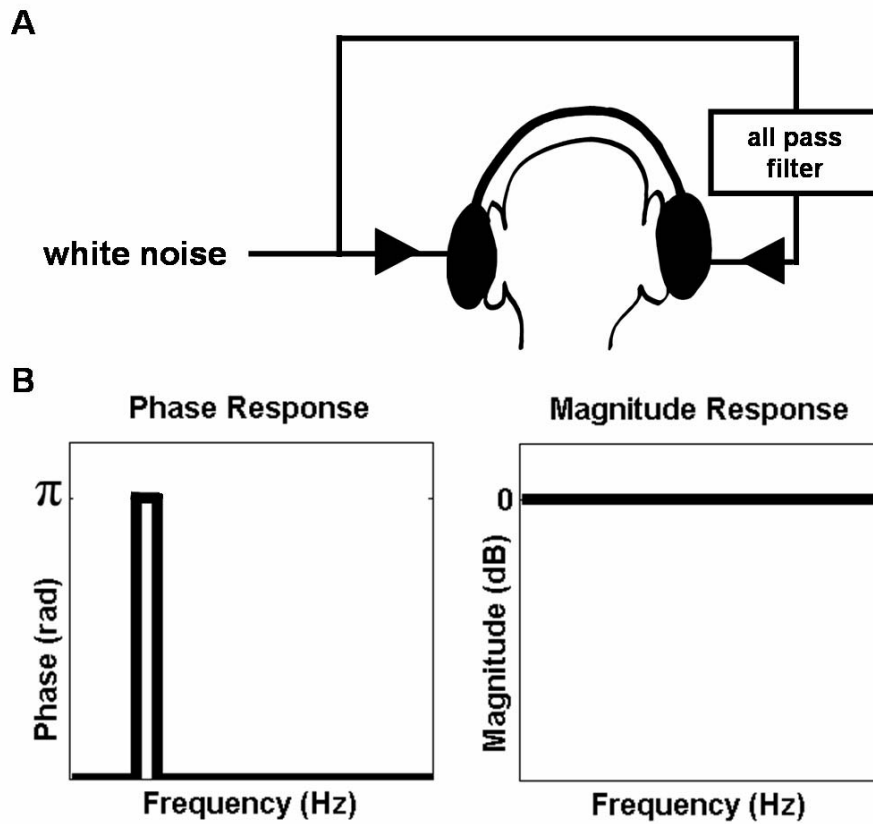
Figure 2.1. Generation of the Huggins Pitch stimuli. (A) The signals were created by introducing a constant phase shift of π in a narrow spectral region of the noise sample delivered to the right ear, while the original sample was delivered to the left ear (note that the particular ear that received the phase shifted noise is of no significance). (B) Schematic of the phase and magnitude responses of the all pass filter. The pitch of the perceived tonal object corresponds to the center of the phase-shifted band.
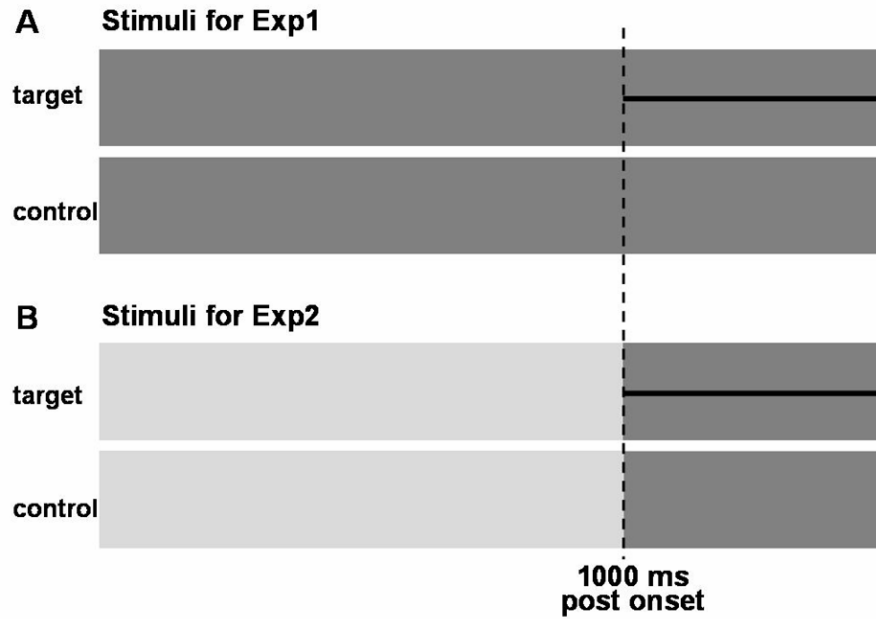
Figure 2.2. Schema of the stimuli used in the two experiments. (A) Stimuli for Exp1 consisted of 1500ms correlated wide-band noise (dark grey) with a 500ms faint tonal object (HP/TN; black line) appearing at 1000ms post onset. Control stimuli were 1500ms long wide-band correlated noise. (B) Stimuli for Exp2 consisted of 1000ms long uncorrelated wide-band noise (light grey)   followed by a 500ms long correlated noise segment which either contained a tonal object (target condition) or did not (control condition). Crucially, the last 500ms of the stimuli of Exp1 and Exp2 were identical.

34

tone-specific component. To attenuate this problem, noise was prepended to the TN and HP signals, so that the tonal response can be measured at the transition from noise to TN or HP, after the stimulus onset response has subsided. Conceptually, the transition response can be seen as evoked by the emergence of a tone-like target within a noise background. In order to investigate the degree to which binaural mechanisms affect the cortical response, the prepended noise was interaurally correlated in Experiment 1 (Exp1), and interaurally uncorrelated in Experiment 2 (Exp2; Fig. 2.2)

The physical differences between HP and TN are explained in Fig. 2.3. Whereas the TN stimuli evoke patterns that can be detected monaurally as early as the auditory nerve (Fig. 2.3B), the HP stimuli are indistinguishable from white noise (Fig 2.3C) up to the level of the Medial Superior Olive (MSO), where phase and timing information from the two ears are first compared (Carr, 1993; Joris et al., 1998). Because HP stimuli are totally devoid of any spectral structure at each ear, they rule out the possibility that their pitch results from spectral processing at the level of the cochlea, auditory nerve or cochlear nucleus. Several studies (Griffith et al. 1998a, 2001; Patterson et al 2002; Krumbholz et al. 2003; Ritter et al., 2005) used a monaural stimulus with characteristics that resemble white noise, iterated rippled noise (IRN), to identify a hypothesized 'pitch center' in Heschl's gyrus (HG) whose activation increased with the degree of temporal regularity in the signal. Noise-like IRN stimuli are intended to reduce the likelihood that the response is related to changes of a tonotopic representation of the stimulus. However this is true only if the iteration frequency is sufficiently low, or the stimulus is high-pass filtered to avoid
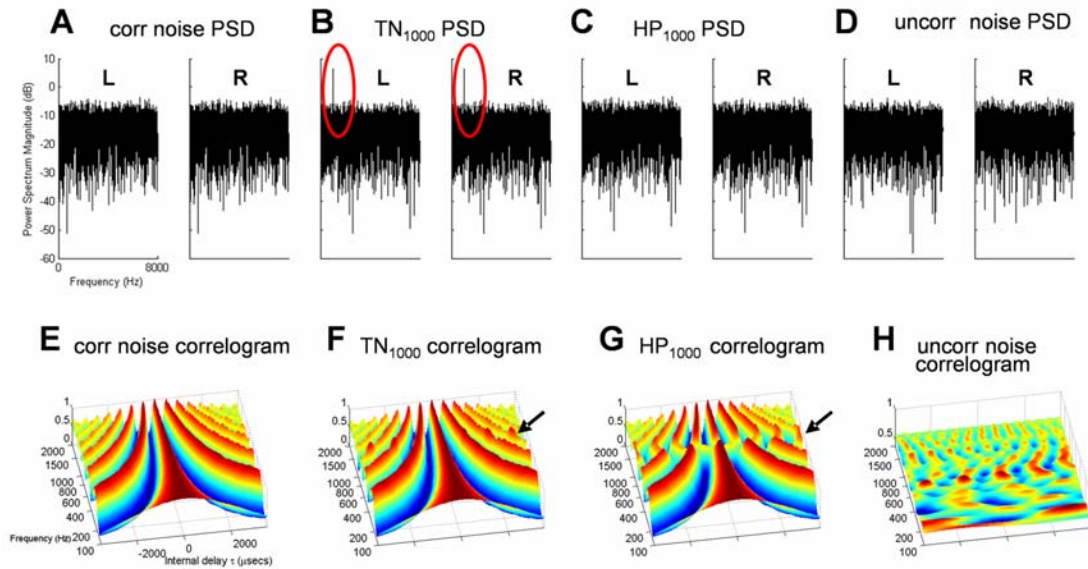
Figure 2.3. Physical properties of HP and TN. Power Spectral Density (L=left ear; R=right ear) was computed for: (A) 1000 ms correlated noise stimulus, (B) 500 ms 1000 Hz TN stimulus, (C) 500 ms 1000 Hz HP stimulus and (D)1000 ms uncorrelated noise stimulus. Pitch information for the TN (but not HP) stimuli is available monaurally at the input to the cochlea (see red circles in B). Physiological evidence indicates that MSO neurons may act as interaural cross-correlators (Joris et al., 1998). Plots (E-H) show binaural cross-correlograms for the stimuli in A-D, which model MSO activation as a neural array arranged by best frequency and best interaural delay (from -3500 to 3500 μs). The plots illustrate the long-term time average of the activity within such an array that would be evoked by our stimuli. The neuronal activation due to correlated noise is shown in (E): peaks at certain delays (main peak at zero ITD with side peaks spaced according to the neuronal best frequency) and troughs at others: Some cells respond strongly to this stimulus (peaks) while others respond weakly (troughs). Activation due to TN1000 (F) is very similar to the correlated noise activation, except for mildly increased activation of already active neurons with best frequency of 1000 Hz (see arrow). In contrast, the HP1000 activation (G) differs sharply from the correlated noise activation - many neurons inactive under correlated noise become active under HP, due to the interaural phase shift in HP (see arrow). The uncorrelated noise stimulus (H) does not activate the MSO as strongly as correlated noise (c.f. Polyakov et al., 1998) and the activation is effectively random. The correlograms were generated using the 'binaural toolbox' (Akeroyd, 2001). The signal is fed through a filter-bank (100 Hz to 2000 Hz with filter spacing 2/ERB) and half wave rectified. Left and right outputs are cross-multiplied and normalized by the average power in the two filter outputs.

resolution of spectral components within the cochlea. HP has no such constraint, and offers the advantage over IRN that spectral cues can be avoided over a range of parameters more typical of salient and musical pitch (Akeroyd et al., 2001).

The requirement for binaural processing before pitch extraction puts constraints on the available mechanisms, as processing can occur no earlier than the site of binaural convergence. By ruling out cues in the periphery, HP stimuli can be used to investigate the generality of the 'pitch center', as proposed by accumulating literature, as well as to refine our interpretation of auditory evoked responses. The M100 or N1 peak (for MEG and EEG, respectively) is the most prominent auditory evoked response. It occurs approximately 100 ms after the onset of a stimulus and is thought to originate from Planum Temporale (PT; Lütkenhöner & Steinsträter,1998). The latency and localization of the M100 have been shown to vary with stimulus parameters that determine pitch (Pantev et al., 1989; Pantev et al., 1996a; Roberts & Poeppel, 1996; Ragot & Lepaul-Ercole, 1996; Fujioka et al., 2003) and it has been hypothesized to reflect stimulus encoding (Salajegheh et al., 2004). Recently, Krumbholz et al. (2003) identified a magnetic deflection (the 'pitch onset response'; POR) which shares some neural generators with the M100 (Seither- Preisler et al., 2004) and is evoked by a transition from noise to iterated rippled noise (IRN). The latency and amplitude of the POR were found to be dependent on the strength and pitch of the IRN stimulus, similarly to findings discussed extensively in the M100 literature. The latency of the M100 onset response to pure tones varies with the tone's frequency such that low frequencies evoke field responses about 30 ms later than high frequencies (Roberts et al., 2000; Lütkenhöner et al., 2001). The period-dependency of latency has been attributed to differences in cochlear traveling wave delays

(Greenberg et al., 1998; Borgmann et al., 2001), or latency differences between neural channels with different characteristic frequencies. By this account HP *should not* exhibit the same trends. Thus comparisons of auditory evoked responses to HP and TN stimuli enable us to test predictions about the architecture of the processing up to and including auditory cortex, and to determine at which point the perceptually similar but physically different stimuli converge on a single representation.

In our experimental paradigm subjects performed a pitch detection task while their brain activity was being recorded. The simultaneous recording of behavioral reaction times and MEG response latencies enables the investigation of the dynamics of the construction of perceptual representations and the degree of correspondence between behavioral and electrophysiological measures. The results reported here have important implications in several domains: in addition to posing new constraints for models of pitch and revealing neural processes associated with the extraction of tonal objects from noise, our data have specific and new implications for binaural processing mechanisms. Importantly, we also demonstrate that, even in humans, electrophysiological tools can measure processes not observable by behavioral, psychophysical means.

2.1.1 Materials and Methods

Subjects

Twenty subjects (mean age 24.6 years), took part in Exp1. Sixteen subjects (mean age 23.8) took part in Exp2. Twelve listeners participated in both experiments. Five subjects (from Exp1; mean age 26.6) took part in the control experiment for Exp1. All

subjects were right handed (Oldfield, 1971), reported normal hearing, and no history of neurological disorder. The experimental procedures were approved by the University of Maryland institutional review board and written informed consent was obtained from each participant. Subjects were paid for their participation.

Stimuli

We chose 4 center frequencies (200, 400, 600, 1000 Hz) that span the frequency region for which HP is salient. A 1000 ms sample of 'frozen' noise was generated for each of these 4 conditions. The signals were created by choosing Gaussian distributed numbers (sampling frequency 16 kHz, bandwidth 8 kHz). The HP signals were generated by introducing a constant phase shift of π in a particular spectral region of the noise sample delivered to the right ear, while the original sample was delivered to the left ear (Yost et al., 1987). The width of the phase-shifted band was set to ±6% of its center frequency (Klein & Hartmann, 1981). The corresponding TN signals were produced by adding a pure tone (with one of the above frequencies) to the same noise samples used to create the HP stimuli. Listeners are able to match the HP signal to the pitch evoked by a pure tone (with a frequency that corresponds to the center of the phase-shifted band) with a standard error of about 3% (Klein & Hartmann, 1981). Three versions of each TN stimulus were created: (1) TN_{center} − perceived in the center of the head (same amplitude of pure tone to both ears); (2) TN_{right} - lateralized to the right (amplitude in the right ear higher than the left ear by approximately 5 dB); (3) TN_{left} - lateralized to the left (amplitude in the left ear higher than the right ear by approximately 5 dB). The amplitude of the pure tone signal was separately adjusted by 2 listeners to match the 'perceived tone' loudness of the corresponding HP

39

stimulus, resulting in a signal to noise ratio (SNR) of approx 10 dB (see Fig. 2.3A-D). The match was verified for each subject in the beginning of the experiment.

The stimuli in Exp1 were 1500 ms long, consisting of 1000 ms interaurally correlated white noise followed by either HP or TN, as described above, or interaurally correlated noise (control). The stimuli of Exp2 were identical to those of Exp1 except that the first 1000 ms of all stimuli were replaced by interaurally uncorrelated noise (Fig. 2.2). The stimuli of the control experiment consisted of 1000 ms interaurally correlated white noise followed by the same noise used to generate the HP stimuli but with one (narrow) band amplified (noise band stimuli, BN). The amplified bands have the same bandwidth as the phase shifted region in the corresponding HP stimulus, but no interaural phase difference. The loudness of the pitch in the BN stimuli was separately adjusted by 2 listeners to match the perceived loudness of the TN stimuli.

When HP is perceived, the background noise is always lateralized to the center of the head but the tonal object may be reported as being at a lateral position away from the midline. It is lateralized to the left or to the right by some listeners but mostly evokes an inconsistent (ambiguous) lateralization, especially by inexperienced listeners (Yost et al., 1987; Zhang & Hartmann, 2004). For the purpose of making the TN and HP stimuli as perceptually similar as possible, prior to the beginning of the MEG experiment proper, each listener's lateralization of HP was assessed. HP stimuli as well as TN_{center}, TN_{left}, and TN_{right} stimuli of the different frequencies were presented in a random order (all stimuli were preceded by correlated or uncorrelated noise, in Exp1 and 2 respectively). Each condition was presented 5 times (total of 80

trials). For each stimulus, the subjects were asked to indicate the perceived location of the tonal object. In cases when subjects were consistent at lateralizing the HP stimuli to the left (3 subjects in Exp1, 5 in Exp2) or right (1 in Exp1, 3 in Exp2) the corresponding TN_{left} or TN_{right} stimuli were chosen for the MEG experiment. If the subjects were inconsistent or indicated that HP was heard in the center, TN_{center} was chosen. Lateralization could also have been obtained by introducing an interaural time difference (ITD), but this would have engaged binaural masking level difference (BMLD) mechanisms similar to those that occur for HP (see discussion). We decided to introduce an interaural level difference (ILD) instead so as to simplify the interpretation of HP/TN differences.

The stimuli were created off-line, gated on and off using 15 ms cosine-squared ramps (with no gating in the transition at 1000 ms post onset), and saved in 16-bit stereo WAV format at a sampling rate of 16 kHz. The signals were delivered to the subjects' ears with a tubephone (E-A-RTONE 3A 50 Ω, Etymotic Research, Inc), attached to E-A-RLINK foam plugs inserted into the ear-canal and presented at approximately 75 dB SPL, to ensure a salient pitch. HP saliency increases with increasing noise level (Durlach, 1962).

In total each subject heard 100 presentations of each of the eight pitch conditions (HP 200, 400, 600, 1000 Hz; TN 200, 400, 600, 1000 Hz) and 800 (50% of all) presentations of the control stimulus. The order of presentations was randomized, with the inter-stimulus interval (ISI) semi-randomized between 500-2000 ms (depending on the subject's RT).

Procedure

Subjects lay supine inside a magnetically shielded room. Before the recording began, each subject's HP lateralization was assessed as described above and the appropriate stimuli were selected. The recording (~1.5 hours) consisted of two parts. First (pre-experiment) subjects listened to 200 repetitions of a 1 kHz 50 ms sinusoidal tone (ISI randomized between 750-1550 ms). These responses were used to verify that signals from auditory cortex had a satisfactory SNR. In the second part of the experiment, subjects, who were not informed about the existence of different types (HP vs. TN) of tonal stimuli, performed a pitch detection task (50% of the trials) by pressing a button, held in the right hand, as soon as they heard a tone popping out of the noise. Reaction times (RT) and accuracy scores were stored and analyzed. Exit interviews showed that subjects were unaware of the existence of different (HP versus TN) tonal stimuli.

For the purposes of (*relative*, i.e. no MR overlay) source localization, five electromagnetic coils were attached to the head of 14 participants in Exp1 prior to the MEG measurement. The locations of the coils were calculated with respect to anatomical landmarks on the scalp using 3D digitizer software (Source Signal Imaging, Inc) and digitizing hardware (Polhemus, Inc). In order to transform the MEG measurements into each participant's individual head coordinate system, the coils were also localized with respect to the MEG sensors. A 3-D head-shape, used to estimate a spherical head model for each participant, was also acquired during digitization.

Neuromagnetic recording and data analysis

The magnetic signals were recorded using a 160-channel, whole-head axial gradiometer system (KIT, Kanazawa, Japan). The data for the pre-experiment were acquired with a sampling rate of 1000 Hz, filtered online between 1 Hz and 58.8 Hz, baseline corrected to the 100 ms pre-onset interval and stored in 500 ms (100 ms pre-onset) stimulus-related epochs. The data for Exp1, 2 and the control experiment were acquired continuously with a sampling rate of 1 kHz, filtered online between 1 and 200 Hz, with a notch at 60 Hz, and stored for later analysis. Effects of environmental magnetic fields were reduced based on several sensors distant from the head using the CALM algorithm (Adachi et al., 2001), and responses were then smoothed by low pass filtering with cutoff at 55 Hz.

In the pre-experiment, auditory evoked responses to the onset of the pure tones were examined and the M100 response was identified for each subject as a dipole-like pattern (i.e. a source/sink pair) in the magnetic field contour plots distributed over the temporal region of each hemisphere. For each subject, the 20 strongest channels at the peak of the M100 (5 in each sink and source, yielding 10 in each hemisphere) were considered to best reflect activity in the auditory cortex and thus chosen for the analysis of the experimental data (Fig. 2.4).

In Exp1 and 2, 1500 ms long epochs (50 ms pre onset) were created for each of the nine stimulus conditions. Epochs with amplitudes larger than 3 pT (~10%) were considered artifactual and discarded. The rest were averaged, low-pass filtered at 20 Hz, and base-line corrected to the full range of the epoch. In each hemisphere, the root mean square (RMS) of the field strength across the 10 channels, selected in

43
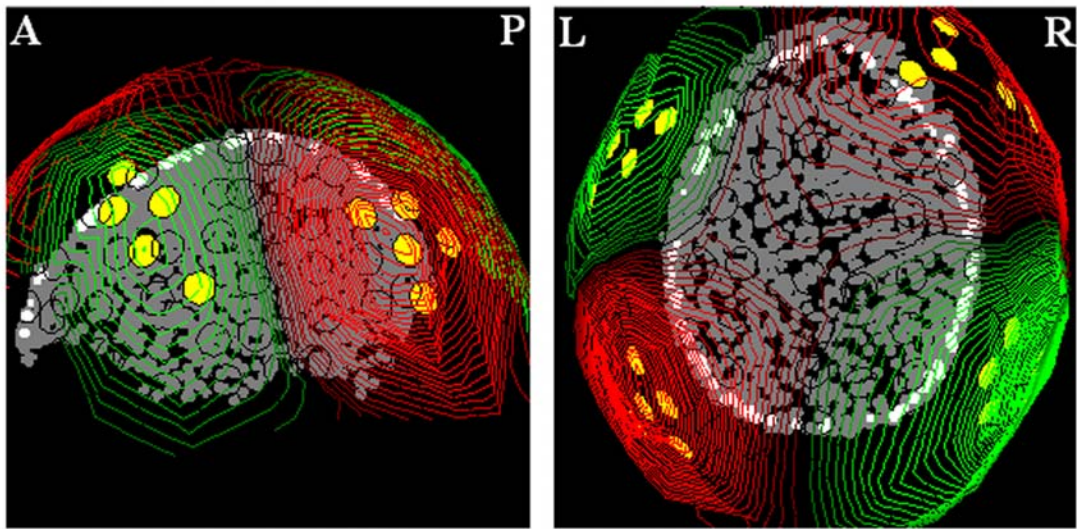
Figure 2.4. Channel selection from the pre-experiment. Different channels were chosen for each individual subject depending on their M100 response; The figure shows a Sagittal View (A=anterior, P=posterior) of the LH and Axial View (L=left, R=right) of the digitized head-shape of a representative subject, the dipole-like pattern in the iso-field maps distributed over the temporal region (red: sink; green: source), and the locations of the 20 chosen channels (yellow circles).

the pre-experiment, was calculated for each sample point. Eighteen RMS time series, one for each condition and each hemisphere, were thus created for each subject. To evaluate congruity across subjects, the individual RMS time series were combined into eighteen group-RMS (RMS of individual RMSs) time series. Consistency of peaks in each group-RMS was automatically assessed with the Bootstrap method (1000 iterations; balanced; Efron & Tibshirani, 1993), a computationally intensive re-sampling method that allows the treatment of situations in which the exact sampling distribution of the statistic of interest is unknown. Source locations were estimated at the RMS peak latency using the model of an equivalent current dipole with the best-fit sphere for each subject's head. A single dipole model was applied for each hemisphere and all channels over that hemisphere were used for the computation.

Since response latencies, which are the major experimental parameter in this study, are naturally characterized by positive skew and the prevalence of outliers, assuming a normal distribution may be misleading. For that reason, for each statistical test presented here, we performed the applicable standard parametric test as well as a form of bootstrapped hypothesis testing (see Efron & Tibshirani, 1993). The two methods yielded very similar results so only the standard parametric test results are reported here. The α level was set a-priori to 0.05. The lower-bound correction was applied where applicable.

2.1.2 Results

The stimuli of Exp1 (Fig. 2.2A) sound like a 1500 ms continuous noise located in the center of the head with a faint tonal object appearing at 1000 ms post onset. The initial portion of the stimuli of Exp2 sounds like a diffused noise (the

binaural stimuli are not fused to a unitary auditory object). At 1000 ms the noise changes from diffused to centered, and at the same time a faint tonal object appears (Fig. 2.2B). This description applies only to binaurally presented stimuli. When listening with only one ear, the stimuli of Exp1 and the corresponding stimuli of Exp2 sound the same (1500 ms of noise in the case of the control stimulus or the HP stimuli and 1500 ms of noise with a tonal object appearing at 1000 ms post onset in the case of TN stimuli).

Experiment 1

Behavioral Data

Subjects were generally accurate at detecting the auditory targets. In Exp1, the mean miss rate was 2% of the pitch trials and the mean false positive count was 2% of the control (noise) trials resulting in d'=4.12. The response time (RT) data are summarized in Fig. 2.5A. An ANOVA with type (HP,TN) and frequency as factors showed significant main effects for both factors ($F(1,19)$=107.456, $p<0.001$; $F(1,19)$=33.167, $p<0.001$) as well as a significant interaction ($F(1,19)$=30.788, $p<0.001$). In Exp1, subjects responded faster to HP stimuli than to TN stimuli, regardless of frequency tested. This effect was significant for all but the 600 Hz stimuli (paired t tests, df=19: 200 Hz, t=-7.35, $p<0.01$; 400 Hz, t=-8.26, $p<0.01$; 1000 Hz, t=-11.83, $p<0.01$).

Electrophysiological data

Waveform analysis reveals that all participants had comparable response trajectories. Fig. 2.6A shows the group-RMS (RMS of the individual-subject RMSs)
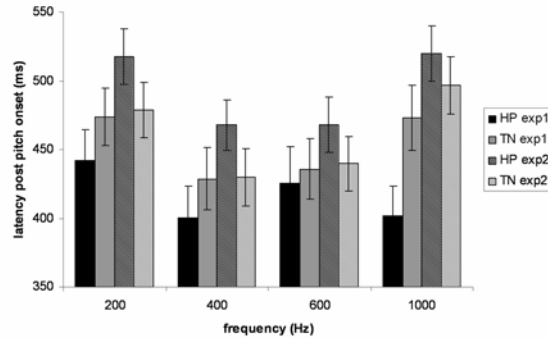
Figure 2.5. Behavioral vs. Electrophysiological responses. (A) Average behavioral RT for the different conditions in Exp1 (solid bars) and Exp2 (striped bars). (B) Electrophysiological peak latency of responses in the LH for the different conditions in Exp1 (solid bars) and Exp2 (striped bars). The time scales are different in the two plots but both show a 200 ms interval to facilitate the visual comparison. (C) Average response time differences (collapsed over frequencies) between Exp2 and Exp1 for electrophysiology and behavior for the 12 subjects common to both experiments. Positive values indicate responses in Exp2 that were delayed relative to Exp1. Electrophysiological responses to TN were earlier in Exp2 than Exp1, opposite to the behavioral pattern and both types of responses to HP. All error bars (in A, B, C) represent 1 std. error.

of the auditory evoked response for each of the conditions in the left and right hemispheres (LH and RH, respectively). The responses contained a two-peaked 'noise onset response' at ~70 ms and ~150 ms post-stimulus-onset (both with a characteristic M50 spatial distribution) and a 'pitch onset response' (POR) (with a characteristic M100 spatial distribution) at ~1160 ms, i.e. roughly 160 ms after onset of TN/HP. The POR (nomenclature introduced by Krumbholz et al., 2003) was modulated in latency by perceived pitch. The dipolar patterns observed in the iso-contour maps for a representative subject are displayed in Fig. 2.6B. Interestingly, the initial 'noise onset response' lacks the usual M100 evoked by stimulus onset. It is likely that the lack of an M100 peak is a consequence of the task performed by the subjects, that led them to attend to the later part of the stimulus and regard the former part as a noise background. This is discussed in detail elsewhere (Chait et al., 2004).

By adding a stretch of noise before the onset of the HP/TN portions of the stimulus, we were able to isolate the brain response to the onset of the stimulus from the response to the onset of the tonal signal. Fig. 2.6A shows that the transient response due to the onset of the noise has faded by around 600 ms. The onset of the pitch corresponds to a prominent increase of activity at approximately 160 ms post pitch onset (~1160 post stimulus onset) and shares important characteristics with the standard M100 response, including its spatial distribution (reflected in the contour plot) and its dependence on perceived pitch (see Roberts et al. 2000 for review). The existence of such a vigorous response is surprising, as it contrasts with the relatively weak perceived loudness of the tonal signals. Repeated measurers ANOVA with hemisphere, type (HP, TN) and frequency (200, 400, 600, or 1000 Hz) as factors
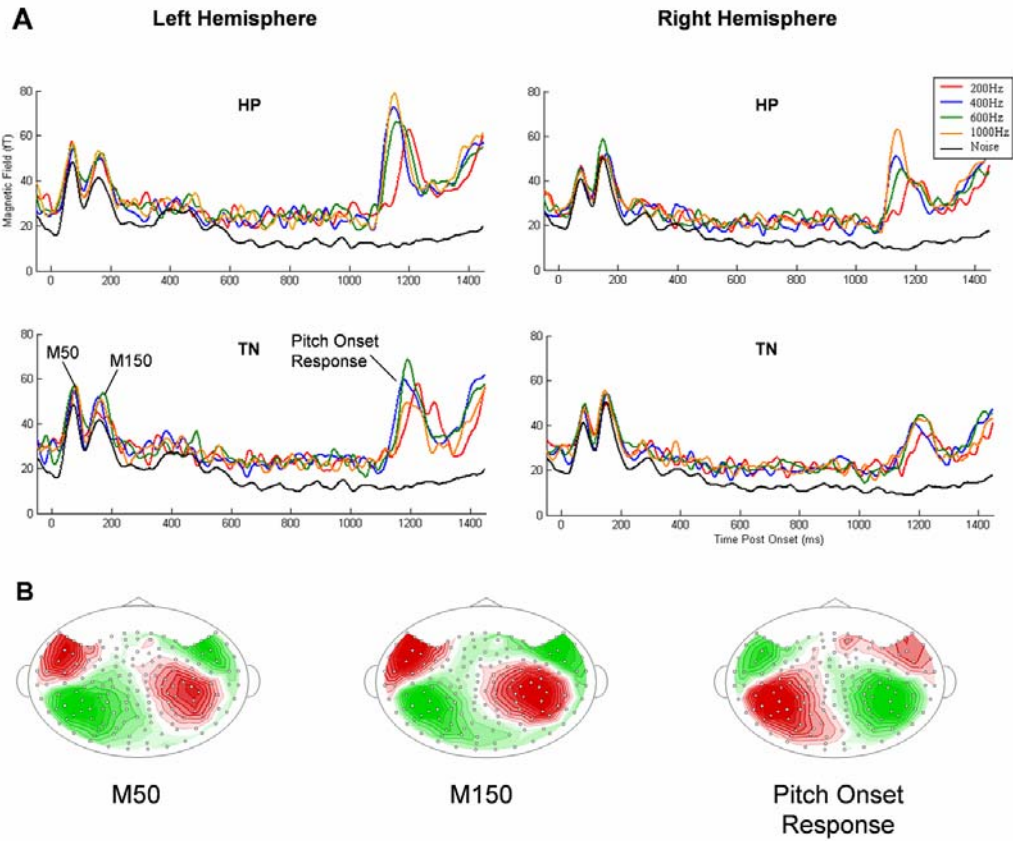
Figure 2.6. Summary of the electrophysiological data from Exp1. (A) The group-RMS in the LH and RH for all tested conditions. The control condition (noise) is lower because it is computed by averaging over many more (800 vs. 100) repetitions. The response is characterized by a two-peaked noise onset response, and a pitch onset response at ~160 ms post HP/TN onset, modulated by frequency. (B) Contour maps from a representative subject at the critical time periods (10 fT/iso-contour). Source=red, Sink=green.
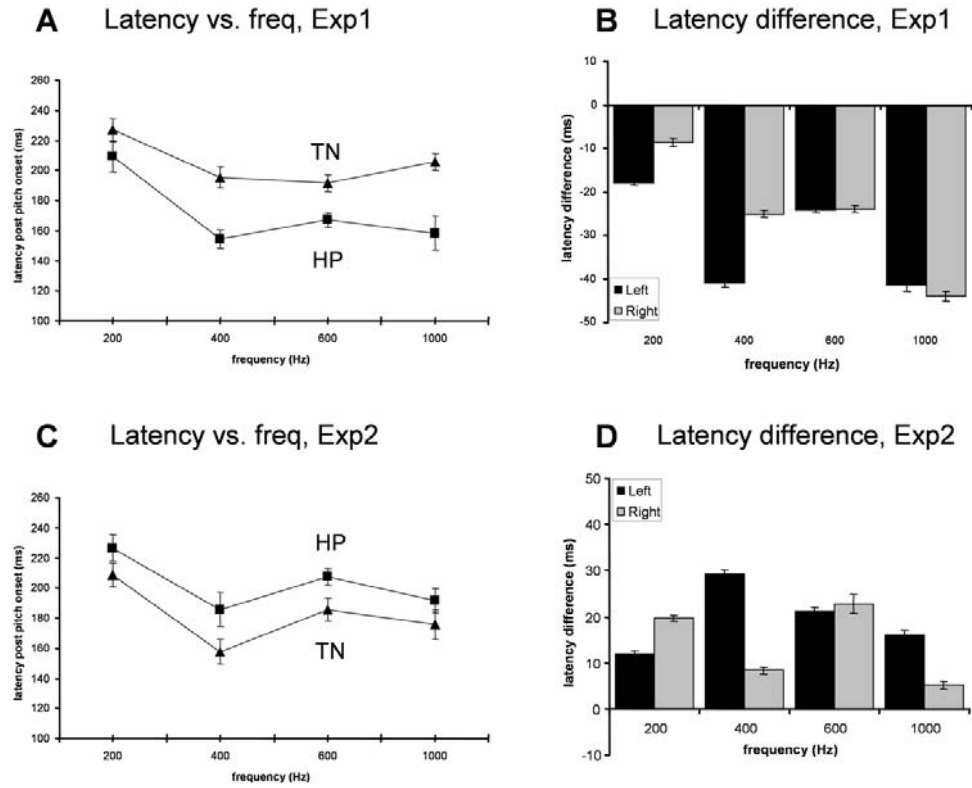
Figure 2.7. Latency results. (A) Exp1: peak latency of the average POR in the LH as a function of frequency. Squares: HP; Triangles: TN. (B) Exp1: Residual latency in the two hemispheres on an individual subject basis. (C) Exp2: peak latency of the average POR in the LH as a function of frequency. Squares: HP; Triangles: TN. (D) Exp2: Residual latency in the two hemispheres on an individual subject basis. All error bars are 1 std. error derived from bootstrap. The values on the x-axis in all figures are category names and are not presented as being on scale.

showed main effects of type ($F(1,19)=65.445$, $p<0.001$) and frequency ($F(1,19)=15.194$ $p<0.001$). The latency of the POR is affected both by the frequency and the type (HP vs. TN) of the signal. HP stimuli elicit a response with a peak latency that is roughly 30 ms earlier than the corresponding TN condition. The observed POR in this study is similar to the POR reported by Krumbholz et al. (2003) in both its spatial distribution and dependency of latency on pitch.

Fig. 2.7A shows the average peak latency in the LH as a function of frequency. The latency of the peak of the POR for each of the 8 conditions was determined automatically for each listener by choosing the maximum value of the RMS in the interval 1100-1300 ms post pitch onset. Note that the peak latency for the lowest frequency is longer, i.e. the 200 Hz signals are associated with significant latency delays, a finding that has been reported and discussed extensively in the context of the M100 literature (Roberts et al., 2000; Lütkenhöner et al., 2001). To estimate the latency differences between TN and HP, for each listener, and each frequency, the latency of the peak of TN was subtracted from that of HP (Fig. 2.7B). The response to HP stimuli is consistently earlier than the response to TN stimuli, 30 ms on average. The amplitude difference between HP and TN peaks was not statistically significant.

The observed latency differences might conceivably be attributed to the difference in bandwidth of the tonal parts of TN and HP stimuli. The EC model of Durlach (1962; 1963), suggests that the internal representation of the HP stimulus resembles that of a narrow band of noise, whereas the TN stimulus is a pure tone. In order to investigate this possibility, a control experiment was run with 5 subjects to

compare the TN stimuli used in Exp1 with noise band stimuli (BN; see methods above). The experimental parameters and procedure were as in Exp1. The data (not shown) demonstrate no significant latency difference between BN and TN for any frequency. In contrast, the same five subjects showed a significant effect in Exp1. Thus, the different activation patterns observed in Exp1 cannot be attributed to a bandwidth difference between the HP and TN stimuli. In the Discussion we argue that they may instead reflect the mechanisms that process binaural stimuli.

A striking finding is that the POR had significantly larger amplitudes in the LH compared to the RH. This effect is found both for HP and TN stimuli (paired t-tests at the peak of the PORs, df=19: TN_{200}, t=3.41, p<0.01; TN_{400}, t=4.16, p<0.01; TN_{600}, t=6.02, p<0.01; TN_{1000}, t=2.66, p=.015; HP_{200}, t=3.81, p<0.01; HP_{400}, t=3.52, p<0.01; HP_{600}, t=5.18, p<0.01; HP_{1000}, t=5.18, p<0.01). This observation is interesting insofar as no such hemispheric differences were found for the M100 response for pure tones in the preexperiment (Fig. 2.8D). Similar findings have been reported (Hertrich et al., 2005; Hautus & Johnson, 2004; Johnson et al., 2003) but are harder to interpret in the latter case because no leading noise or tone controls were used. Krumbholz et al. (2003) measured MEG signals only over the left hemisphere.

The stimulus onset responses also showed significant hemispheric differences (discussed in Chait et al., 2004). Hemispheric lateralization thus switched (at least) three times during the full course of the 1500 ms stimulus (Fig. 2.8). The first peak, M50, is stronger in the LH, the second one, M150, is stronger on the RH, and the third peak, POR (the 400 Hz stimuli are shown as an example), is again significantly larger in the LH.
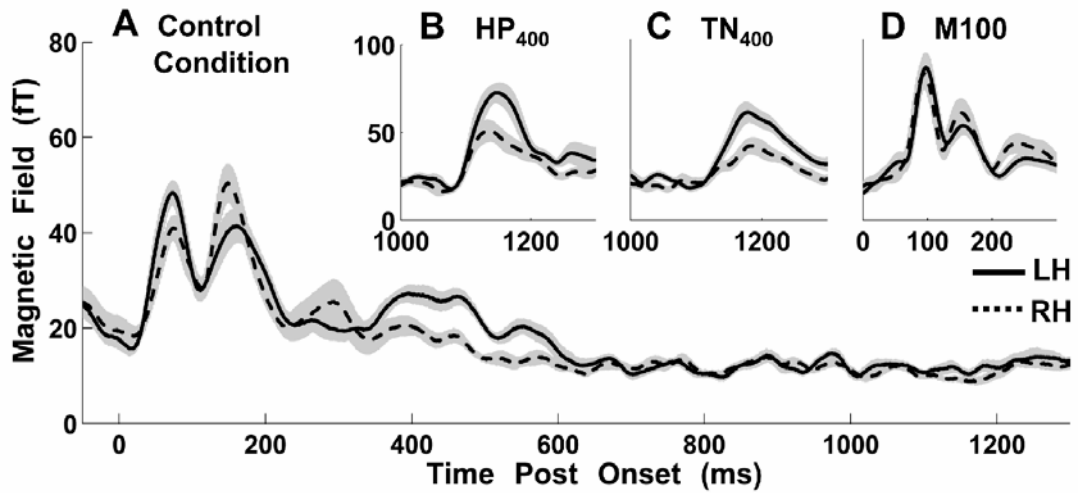
Figure 2.8. Comparison of hemispheric activation: LH, solid lines; RH, dashed lines. Grey shadings are 1 std. error derived by bootstrap. (A) Control condition in Exp1. M50 peak has stronger activation in the LH; M150 peak shows stronger activation in the RH. (B, C) POR for all HP and TN conditions (400 Hz shown here as an example) exhibited stronger left hemispheric activation. (D) The M100 response in the pre-experiment (1 kHz tones) showed no significant hemispheric differences.

As discussed in the methods section, we compensated for listener-specific subjective lateralization of HP by using TN stimuli with similar perceived lateralization as HP (see methods). In all of the findings discussed above, there was no difference between the subjects who perceived HP on the left or right (and thus were presented with TN_{left} or TN_{right} in the experiment) and the subjects who heard HP at the center of the head or at an ambiguous location (and thus were presented with TN_{center}). Therefore, the hemispheric lateralization and the latency differences of the MEG response do not seem to be related to the perceived location of the stimuli. Additionally, in experiments specifically designed to investigate spatial (interaural time difference; ITD) effects on auditory evoked potentials the observed latency effects were on the scale of a few milliseconds (McEvoy et al., 1993), an order of magnitude smaller than those observed here.

The source of the Pitch Onset Response

For each of the 14 subjects for whom a digitized head shape was obtained, the M50 peak (mean latency=69.8 ms) in the control condition in the LH was fitted to a single equivalent current dipole. One listener with a goodness of fit (GOF) of less than 80% was discarded from the analysis. The mean GOF for the 13 remaining listeners was 88.77%. The M50 component is believed to originate in or near PAC. In a recent study, Yvert et al. (2001) found it to activate the antero-lateral portion of HG and Heschl's sulcus. This might reflect activity in the human counterpart of the anterior areas in the core line region or in the antero-lateral belt region described in monkey (Kaas et al., 1999).

In order to compute the proportion of the POR field explained by the current dipole obtained for the M50 components, we estimated the GOF of that M50 dipole (maintaining a fixed location and orientation but allowing for a 180 degree flip in polarity) for the pitch onset component. The 400-Hz stimuli, HP$_{400}$ (mean latency = 1135 ms) and TN$_{400}$ (mean latency = 1186 ms), were characterized by the clearest and most prominent POR peaks and thus were chosen for this analysis (for the same reason only the LH response is fit). The resulting mean GOF was 77.3% for HP$_{400}$ and 76% for TN$_{400}$, with no significant difference between conditions. These findings indicate that the M50 dipole produces a good fit to the field of the POR. The M50 and the POR have opposite polarities, so the two processes cannot originate from identical neuronal populations. The good negative fit suggests at least one possible explanation: that the sources of the activity lie in close proximity in auditory cortex, though perhaps on opposite sides of a cortical fold. When looking at the proportion of the field explained by the M50 dipole in the time range 0-1400 ms, the time windows with the highest GOF are around the onset of the stimulus (M50 time window) and the onset of the pitch. The average GOF of the remainder, including, crucially, the time window around 100 and 150 ms post noise onset, is below 40%, suggesting a different source and further affirming the specificity of the relation between the M50 component and the POR. These findings (that the POR originates in an area not in PAC but very close to it) are consistent with recent fMRI studies: Penagos et al. (2004) report that a region immediately anterolateral to PAC exhibits sensitivity to pitch salience. In Patterson et al (2002), the contrast between noise and pitch eliciting iterated rippled noise activated a relatively small, bilateral region, lateral to PAC on

HG, which the authors suggest might correspond to the R or RT region of core auditory cortex.

The small number of repetitions per condition, in combination with the spatial resolution of the MEG system and inter-subject variability prevent us from making any further spatial conclusions with a reasonable degree of certainty.

Experiment 2

The only respect in which Exp2 differed from Exp1 was that the initial 1000 ms of all stimuli (including controls) were *uncorrelated* instead of *correlated* noise. Crucially, the last 500 ms (HP, TN, or noise for the control stimuli) remained exactly the same as in Exp 1. In particular, the noise in these segments was *correlated*, and thus switched from uncorrelated to correlated at 1000 ms post onset (see Fig. 2.2).

Behavioral Data

Performance in Exp2 was slightly worse than in Exp1. Those subjects that participated in both experiments (N=12) observed that Exp2 was more difficult. They reported that in addition to the change in the noise that occurred simultaneously with the appearance of pitch and hindered the detection, the quality of the noise (as two unfused objects at each ear) in the first 1000 ms of the stimulus made it harder to detect the tonal object. This was reflected in the moderately increased error rate in Exp2: the mean miss rate was 7% of the tonal trials and the mean false positives count was 4% of the control (noise) trials resulting in $d'=3.23$.

A repeated measures ANOVA with type (HP, TN) and frequency as factors

revealed main effects of both ($F(1,15)=29.044$, $p<0.001$; $F(1,15)=52.77$, $p<0.001$). In contrast to Exp1, in Exp2 subjects responded faster to TN stimuli than to HP stimuli (paired t tests, df=15: 200 Hz, t=5.45, $p<0.01$; 400 Hz, t=6.13, $p<0.01$; 600 Hz, t=4.11, p=0.01; 1000 Hz, t=2.12, p=0.051; Fig. 2.5A);

Electrophysiological Data

Fig. 2.9A shows the group-RMS of the auditory evoked response for each of the conditions in the LH and RH. The dipolar pattern in the iso-contour maps for a representative subject is displayed in Fig. 2.9B. The onset of the stimulus is characterized by two peaks, at 70 ms and 180 ms in the LH, and at 70 ms and 170 ms in the RH. As in Exp1, the classical M100 response is absent at the onset of the noise (Chait et al, 2004).

In contrast to Exp1, in Exp2, the separation of the response to the onset of the tonal target from the response to the noise is not as complete because there are two simultaneous changes occurring in the stimuli at 1000 ms post onset: (1) the noise changes from uncorrelated to correlated and (2) onset of HP/TN (see Fig 2.2). However, comparison of responses to the control stimulus and pitch-evoking stimuli indicates that these two changes are processed at different times: The cortical activity due to the change in the noise is reflected in a peak in the control condition at 1134 ms in the LH and 1130 ms in the RH. The latency of this peak was not modulated by the perceived pitch of the tonal stimulus (when present). The POR is evident in the pitch conditions (especially in the LH) as a separate, second peak at approximately 1160 ms post onset or later, so the two responses can be distinguished. A repeated measures ANOVA with hemisphere, type (HP, TN) and frequency as factors showed
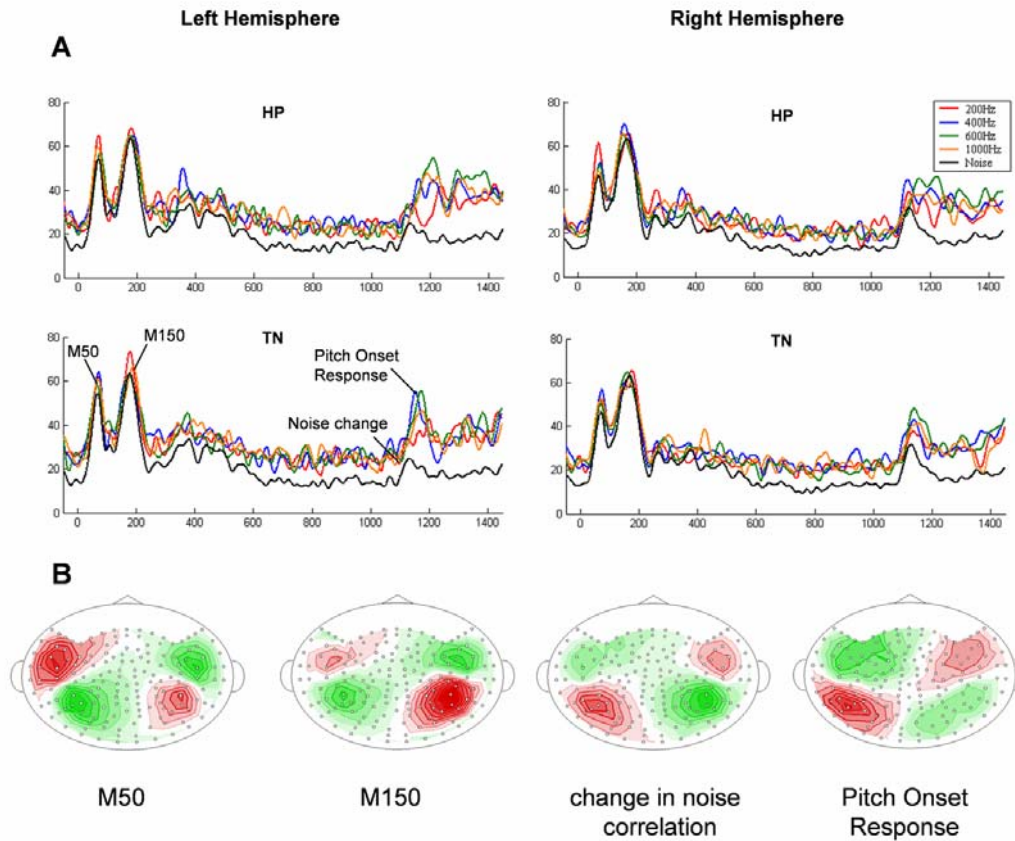
Figure 2.9. Summary of the electrophysiological data from Exp2. (A) The group-RMS of in the LH and RH for all tested conditions. The control condition (noise) is lower because it is computed by averaging over many more (800 vs. 100) repetitions. The responses contain a two peaked noise onset response, a peak that corresponds to the change in noise at ~1130 ms post onset, and a pitch onset response at ~160 ms post HP/TN onset, modulated by perceived frequency. (B) Contour maps

main effects of type ($F(1,14)$=11.8 p=0.004) and frequency ($F(1,14)$=2.88, p=0.047). As in Exp1, the latency of this peak covaries with perceived pitch and type: HP stimuli elicit a response with a peak latency that is roughly 20 ms *later* than the corresponding TN condition. The difference between the amplitude of the peaks of HP and corresponding TN stimuli is not significant.

Fig. 2.7C shows the average peak latency in the LH as a function of the frequency of the perceived tonal object. The latency of the peak of the POR for each of the 8 conditions was automatically determined for each listener by choosing the maximum value of the RMS in the interval 1100-1300 ms post pitch onset. The data in the 200-Hz condition were noisy and peaks were not found in the data of 2 subjects. For 5 additional participants, the maximum value of the RMS was at the peak associated with the change in noise, in those cases the POR was defined as the second highest peak in the above specified interval. In order to estimate the latency differences, for each listener and each frequency the latency of the peak of TN was subtracted from that of HP (Fig. 2.7D). Positive values reveal that the response to HP stimuli is consistently *later* than the response to TN stimuli by approximately 20 ms.

The physiological data from Exp2 were noisier than the data from Exp1. However, the main difference with Exp1, the switch in latency between TN and HP stimuli, is robust across frequencies and listeners (see Fig 2.7) and cannot be attributed solely to a decrease in SNR.

Hemispheric comparisons do not yield significant differences in Exp2, possibly as a consequence of the noisier nature of the data, but the trend is in the same direction as Exp1. Interestingly, the peak at approximately 1130 ms, associated with

the change in the perception of the noise, had higher amplitude in the RH (not statistically significant, but approaching significance). This might be related to fMRI reports of stronger activation in the RH when subjects were listening to changes in binaural timing (Griffiths et al., 1998b). The higher amplitude in the RH is probably masking more of the pitch onset signal, which in turn might explain the noisier results in the RH (see Fig. 2.9).

As in Exp1, the hemispheric lateralization or the latency differences between the stimuli were not related to the perceived location of the stimuli.

Comparison between Experiment 1 and Experiment 2

Behavioral data

When comparing the results of the two experiments (three-way repeated measured ANOVA for the 12 subjects who participated in both experiments; corrected) there is a main effect of experiment ($F(1,11)$=13.33 p<0.01) and frequency ($F(1,11)$=45.61 p<0.01) as well as an interaction of experiment×frequency ($F(1,11)$=12.12 p<0.01), stimulus type×frequency ($F(1,11)$=11.91 p<0.01), experiment×stimulus type (HP/TN $F(1)$=48.15 p<0.01), and experiment×stimulus type×frequency ($F(1,11)$=9.96 p<0.01), indicating that RTs for both HP and TN were longer in Exp2 than in Exp1. Despite the three-way interaction, effects at all frequencies are similar. When collapsed across frequencies (fig 2.5C) a paired sample t-test showed that RTs for both HP and TN stimuli were greater in Exp2 relative to Exp1 (df=47; HP: t=-9.837, p<0.01; TN: t=-2.261, p<0.028), corresponding to subjects' reports that Exp2 was harder; individual frequency results are similar but

noisier.

Electrophysiological Data

We compared POR peak latencies of the two experiments using a three-way repeated measures ANOVA (with experiment, hemisphere, type and frequency as factors) for the 12 subjects that participated in both experiments. Unlike for the behavioral data, there were no main effects of experiment or stimulus type (due to the latency reversal in the two experiments) but there is a main effect of frequency ($F(1,10)$=5.82, p=0.037), as well as an interaction between experiment × stimulus type (HP/TN) ($F(1,10)$=34.73 p<0.01) .

Fig. 2.5B summarizes the latency data of the PORs from the two experiments (we report results for the LH; RH results are similar but noisier). When collapsed across frequencies (fig 2.5C) a paired sample t-test showed a significantly greater latency in Exp 2 relative to Exp 1 for HP stimuli, but *smaller* latency to TN stimuli (df=47; HP: t=-2.789, p<0.01; TN: t=3.17, p<0.01); The amplitude of the POR in the LH did not show a significant change between the stimuli of Exp1 and those of Exp2. We argue in the discussion that the different peak latencies of responses to HP and TN in the two experiments may be attributable to the different activation that they produce in the binaural system.

When performing the pitch detection task, subjects pressed a button held in their right hand. It could be argued that this might have influenced the amplitude and left lateralization of the pitch onset peak. Such an effect can be dismissed by comparing the data from the two experiments: the experimental procedure was identical in both experiments, accuracy of pressing the button was comparable (see

above), but the latency, amplitude and lateralization of the peaks differed significantly. This suggests that these characteristics of the responses reflect the processing of the acoustic stimuli and not the motor event.

2.1.3 Discussion

The primary objective of this whole-head auditory MEG study was to investigate the timing of the formation of the percept of tone in two physically very different signals that elicit a very similar pitch percept – Huggins Pitch (HP) and a pure tone in noise (TN). Several aspects of the data, as well as new unresolved issues, are discussed in turn below.

2.1.3.1 Implications for Pitch

In this study we observed a prominent cortical response to the onset of both HP and TN stimuli, occurring at approximately 150-200 ms post tonal onset. By prepending noise to the HP/TN stimuli we are able to separate the response to the tonal onset from other processing associated with the onset of an acoustic stimulus. Unlike other studies that used the M100 response at the onset of a tonal stimulus to probe brain processes that handle pitch, we measured the response to the transition from a noise to a pitch-like stimulus embedded in noise (as in Krumbholz et al, 2003). The latency of the M100 response at the onset of a tonal stimulus is known to covary with pitch, but this response confounds stimulus-onset-related computation with pitch computation. In contrast, the response to the transition from noise to a pitch-like stimulus (POR) allows pitch computation to be isolated. Additional evidence for the

fact that the POR truly reflects tonal processing and not just general change in the input come from Exp2 where the deflection corresponding to the change in noise from uncorrelated to correlated occurs approximately 30 ms earlier than the response to the onset of the tonal targets, even though both changes occurred simultaneously in the signal (see Fig. 2.2). If the POR were just a detector of change in the ongoing stimulation (e.g. Jones et al., 1991), we would expect to see a single peak rather than the two peaks observed in Exp 2, each with distinct temporal and spatial properties: POR stronger on the LH, response to change in noise stronger on the RH, indicating distinct neural generators.

The POR peaked at 160 ms post tonal onset, approximately 60 ms later than a classic M100, although M100 latencies to near-threshold tonal signals are comparably delayed (Stufflebeam et al., 1998). The early sensory processing of HP differs from TN and the other diotic stimuli that were investigated in previous experiments (regular interval noise in Patterson et al 2002, Krumholtz et al 2003 and Ritter et al., 2005; stimuli with and without resolved harmonic in Penagos et al 2004). The comparison of the activation evoked by HP and TN stimuli therefore affords an opportunity to distinguish neural processing common to these pitch-like stimuli from those specific to early sensory processing. The responses to both (TN/HP) were localized to the same cortical area and exhibited similar frequency dependence. For both, low frequency stimuli elicited longer latencies than high frequency ones, similar to the 'classic' M100 response (Roberts et al, 2000). Since HP showed this effect, an inevitable conclusion is that the dependence on frequency originates from central mechanisms (beyond the stage where binaural information is combined); explanations

based solely on the cochlear traveling wave delay (Greenberg et al., 1998; see also Borgmann et al., 2001) must therefore be reassessed. Crucially, this data indicates that by approximately 150 ms post onset both types of tonal objects are mapped to very similar representations in cortex. The similarity of the MEG responses is consistent with behavioral data indicating that Huggins Pitch generates pitch and timbre percepts that are like those of monaural tones (Bilsen, 1977) and suggests that these signals are processed similarly, despite their physical differences.

Our data provide new evidence that the processes that give rise to the POR reflect *central* pitch mechanisms. Patterson et al. (fMRI; 2002) and Krumbholz et al. (MEG; 2003), using monaural iterated rippled noise (IRN) stimuli, identified a hypothesized 'pitch center' in Heschl's gyrus (HG), whose activation increased with the degree of temporal regularity in the signal. Since the pitch-evoking structure of IRN is present in the stimulus at both ears, its processing could begin as early as the cochlear nucleus (as suggested by Griffiths et al., 2001). Such is not the case for HP. According to generic models of auditory processing (McFadden, 1975; Colburn & Durlach, 1978; Stern & Trahiotis, 1995), auditory information is processed via monaural and binaural analyzers. Huggins pitch requires binaural presentation, and thus cannot be extracted by low-level processors within a monaural pathway that precedes binaural interaction. The pitch of the TN stimulus (like IRN) can, in theory, be detected monaurally, and consequently could be extracted peripherally or at some low-level stage (cf. Cariani & Delgutte, 1996). Parsimony, and the fact that TN varies with frequency similarly to HP suggests that *both* stimuli are processed beyond the stage of binaural convergence, and argues against the hypothesis of separate pathways

for monaural and binaural pitch phenomena. Different latencies for TN and HP responses might be taken as evidence for separate processing levels, but the fact that HP MEG responses *preceded* TN responses in Exp 1 would imply processing of the binaural HP stimulus occurs at an earlier stage than that of the monaural TN stimulus, which seems unlikely. Instead, we propose the latency differences arise as a result of binaural processing that both signals undergo (see below).

Interestingly Carlyon et al. (2001), using dichotic pulse trains with no place information, were led to conclude that temporal pitch mechanisms operate on the input to each ear alone rather than on the output of the binaural system (combination of the information from the two ears). That conclusion was obtained using unresolved stimuli. Our study suggests the opposite conclusion (not only for HP, but also for TN), but as we used resolved stimuli (in the sense that the tonal components are isolated sinusoids) we do not know whether our conclusion applies to spectral (place) or temporal pitch mechanisms or both.

Our results are consistent with a pitch processor that is driven by a central spectrum, computed from all available information irrespective of it being monaural or binaural in nature. The EC model of Durlach (1962; 1963), for example, suggests that the central representation of HP has a similar time and place profile to that of TN. In this study we have only used pure tones (TN) and 'pure tone'-like stimuli (HP). Nevertheless the data reported here are a significant first step and constitute a generalizable paradigm for use with richer pitch stimuli. An intriguing next step is comparing brain responses to complex tones with a missing fundamental with response to missing fundamental HP complexes (Bilsen, 1977). The 'Bilsen multiple

phase shift pitch', has phase shifts at several harmonics, and therefore is the counterpart of the pitch of harmonic complex tones. HP and its generalizations are ideal tools to study 'missing fundamental' effects because combination-tones resulting from cochlear non-linearities (Pantev, 1989) do not exist.

The LH advantage at the peak of the POR is unexpected. No hemispheric difference is reported in studies involving detection or discrimination of pure tones (Papanicolaou et al., 1999; Shtyrov et al., 2000; Johnsrude et al., 2000; but see Devlin et al., 2003), or in studies with IRN stimuli (Patterson et al., 2002; though see Griffiths et al., 1998a). In fact, pitch related tasks are usually reported to produce stronger responses from the RH (Zatorre, 2001). One possible explanation is that the pitch onset peak does not reflect the extraction of pitch per se, but the segregation of the tonal object within the auditory scene and its separation from the noise background, which might incur more significant involvement of LH mechanisms. The onset of the pitch-evoking stimulus activates pitch sensitive mechanisms that produce a pitch-like percept, but also segregation (or pop-out-of-background) mechanisms that produce the perception of a pitched *object*. The POR may be tapping into the latter stage. Alain et al. (2002; see also Dyson & Alain 2004), in EEG studies of concurrent sound segregation, reported that the perception of a mistuned harmonic as a separate sound is associated with a negative wave peaking at about 150 ms after sound onset. In their stimuli the onset of mistuning coincides with that of the stimulus, so the response components of the two cannot be isolated, but the properties of the wave (referred to as object-related-negativity) are very similar to those of the POR (see also Hautus & Johnson, 2004). This alternative interpretation, which could equally be

applied to the results of the other fMRI/MEG studies cited here (e.g. Patterson et al., 2002; Penagos et al., 2004, Pantev et al., 1996a; Fujioka et al., 2003; Krumbholz et al., 2003), still allows the responses to be interpreted in terms of pitch processing. Extraction of the pitch 'feature' must precede object emergence, and indeed the cortical response latency showed a dependence on the pitch of the tonal object. This study cannot disambiguate between an interpretation in terms of a pitch-specific response and this alternative. A more systematic and detailed examination of this question is warranted.

2.1.3.2 Binaural processing

Models of auditory processing are often discussed in terms of 'monaural' and 'binaural' pathways (McFadden, 1975; Colburn & Durlach, 1978). The former are invoked to account for phenomena involving monaural (or diotic) stimulation, the latter for phenomena that arise only from dichotic stimulation. The pitch evoked by HP must arise in the latter, but that evoked by TN might conceivably arise in a monaural pathway. We do not see evidence of such a division in the present study. As discussed above, when presented monaurally the stimuli of Exp1 and Exp2 are indistinguishable; consequently any differences between the two experiments must be due to binaural mechanisms. In particular, the latency differences observed between Exp1 and Exp2 for TN (Fig. 2.5C) must result from the binaural processing of correlated vs uncorrelated noise. The change in the interaural configuration of the leading noise caused an *earlier* brain response to TN in Exp2 relative to Exp1. Note that a difference of the opposite sign would have been compatible with the hypothesis

of monaural processing of the pitch of TN with subsequent interference from the result of binaural processing at a later stage. The fact that the responses in Exp2 were *earlier* indicates that binaural processing *aided* the detection of TN, and thus that detection occurred within a binaural rather than monaural pathway. HP is commonly hypothesized to be mediated by the mechanism of binaural unmasking (Raatgever & Bilsen 1986; Culling et al., 1998): A target that is just masked by binaurally correlated noise can be made easier to detect by inverting the noise or the target in one ear. Unmasking depends on mechanisms that are sensitive to the similarity of the signals at the two ears. Cells exhibiting these properties are found in the MSO, and in animal studies, the Inferior Colliculus (IC), the projection target of the MSO, exhibits correlates of binaural unmasking (Jiang et al., 1997ab; Palmer et al., 2000).

The auditory evoked magnetic fields, measured outside the head by MEG, are generated by neuronal currents flowing in tens of thousands of cortical pyramidal cells on the supratemporal gyrus (Hämäläinen et al., 1993), but the observed response latency differences might originate as early as the SOC. One possibility is that the latency disparity reflects constraints of processing within the MSO itself. The MSO has been likened to an array of cross-correlators fed from both ears (Jeffress, 1948; Joris et al., 1998). Fig. 2.3 (E to H) illustrates the long-term time average of the activity within such an array that would be evoked by our stimuli. Correlated noise (Fig. 2.3E) evokes an orderly arrangement of 'valleys' and 'peaks' (resulting from each cell's sensitivity to a particular relative phase between the two ears at its best frequency; c.f. Yin & Chan, 1990). TN stimuli (Fig. 2.3F) evoke the same pattern with slightly higher amplitude at the frequency of the tone, whereas HP stimuli (Fig.

2.3G) produce a more complex pattern with a crossover between ridges at the frequency of the phase transition. Uncorrelated noise (Fig. 2.3H) evokes an irregular pattern with low amplitude. The frequency-local features of target stimuli (TN or HP) distinguish them from non-target correlated noise. The influence of the preceding context (correlated or uncorrelated noise) on their detection might explain the latency differences observed between Exp1 and Exp2. In Exp1, the onset from correlated noise to HP activates neurons, within the frequency region of the phase transition, that were previously inactive (compare Figs 2.3E and 2.3G). In Exp2 the onset from uncorrelated noise to HP also increments activity, but in this case it is distributed across the frequency axis. Easier detection of tonal HP targets in the former case might explain the smaller latencies of brain responses observed in Exp1. For TN, in Exp1 the onset of the tonal target causes a local increment in the activity of neurons that were already strongly activated by the correlated noise (compare Figs. 2.3E and 2.3F). In Exp2 the same neurons were less strongly activated by the uncorrelated noise (compare Figs. 2.3F and 2.3G). Easier detection of tonal TN targets in the latter case might explain the smaller latencies observed in Exp2 for brain (but not behavioral) responses. The patterns illustrated in Fig.2.3 reflect the generic cross-correlation model of Jeffress, but a similar account could be applied to the recent model of McAlpine and colleagues (McAlpine & Grothe, 2003). Thus, latency differences may result from constraints of binaural processing as early as the MSO, but it is not clear how they would result in the relatively large latency differences observed. It is also not clear whether they arise at the MSO itself, or in subsequent stages that interpret its output. To the best of our knowledge, no physiological studies

have tried to measure the latency of responses to such stimulus events at the level of MSO or IC. Several different models of unmasking (and HP) have been suggested to account for the available psychophysical, physiological and electrophysiological data (Durlach, 1963; Raatgever & Bilsen 1986). Physiological measurement of the latency of responses to signals such as those used here might greatly clarify our understanding of the processing of these events.

2.1.3.3 Behavior vs. Electrophysiology

Although behavioral and electrophysiological responses mostly follow similar directions, there is also a striking difference between them (Fig. 2.5C). In the behavioral data, the average RT for HP was significantly greater in Exp2 than in Exp1. Similarly RTs for TN were either equal or greater in Exp2 relative to Exp 1. In contrast, in the electrophysiological data responses to TN stimuli are *faster* in Exp2 than Exp1 (negative values in fig 2.5C), while the responses to HP stimuli are slower in Exp2 than Exp1 (positive values in fig 2.5C). Thus, although HP responses follow the same pattern as behavior, brain responses to TN in Exp2 (where changing binaural cues are present) are *earlier* than those to TN in Exp1 (no changing binaural cues). The difference between behavior and brain response patterns can be understood by supposing that the transition from correlated to uncorrelated background in Exp2 introduced an additional difficulty (reported by the subjects) that caused RTs to be overall longer in Exp2 than in Exp1.The fact that it was not apparent in the POR brain responses suggests that it affected a different pathway (or a stage subsequent) to that which produced those responses.

Behavior and electrophysiology, studied separately, might lead to different conclusions about the nature of the processing involved. The simultaneous acquisition of both MEG and behavioral data puts stronger constraints on the interpretation, revealing a multi staged process where early (approximately 150ms post onset) cortical responses (POR) reflect the operation of low-level mechanisms but behavior is affected by additional mechanisms. For example, these mechanisms might incorporate the outputs from the POR generating system as well as the outputs from a separate system sensitive to the change in the background noise into some decision variable, in this manner making the conscious detection of the tone onsets in Exp2 slower than in Exp1.

The incongruence between behavior and electrophysiology observed here demonstrates that there is a limitation on what can be learned from behavioral or electrophysiological measure alone. In the study of the processes that underlie the construction of perceptual experiences, electrophysiological measures can usefully supplement the wealth of data that have accumulated over the (relatively) long history of behavioral research.

2.2 The dynamics of the construction of perceptual representations[3]

Sound sources within the environment produce an aggregate wave-form that enters each ear. To make sense of the world, a listener must separate this input into source-related components, localize them, recognize them, and react accordingly. This section concerns the processes by which auditory *sensory* information is converted into a *perceptual representation relevant for behavior*. These processes are mostly automatic and do not require explicit effort (e.g. Goldstein, 2005; Wolfe et al., 2006): We follow a conversation in a noisy restaurant, or close our eyes in the park and still perceive the birds on the tree, the dog barking in the corner and the children playing in the playground. 'The essence of perception' has engaged human interest throughout scientific history. These processes have been studied through introspection (e.g. Aristotle's On the soul in (Barnes, 1984), (Schwartz, 2004)) and later with psychophysics (Boring, 1942; Yost et al., 1993; Moore, 1997). Careful behavioral experimentation is instrumental in penetrating the private nature of perception (which is, by definition, an *internal* representation) and measuring what it is a listener perceives: What is the perceptual loudness of the stimulus (Green, 1993)? What is its perceived pitch (Moore, 1993)? Which parts of the stimulus are assigned to foreground and which ones are relegated to background (Brochard et al., 1999)? This is especially important since we know that identical physical inputs may result in

---

[3] This section is based on Chait M, Simon JZ (in press) The dynamics of the construction of auditory perceptual representations. In: Reasoning and Cognition Vol 2, Tokyo: Keio University Press.

different percepts, depending on the state of the listener (Bregman, 1990; Chait et al., 2004; Micheyl et al., 2005).

Psychophysical experimentation has also played an important role in formulating models of the underlying brain systems and computations leading to perception. For example, by measuring listeners' sensitivity to different sounds, researchers have been able to put constraints on pitch processing mechanisms (e.g. Oxenham et al., 2004; de Cheveigne, 2005), the structure of the peripheral auditory system (e.g masking experiments by Fletcher, 1940), binaural processing mechanisms (Blauert, 1997) and the features into which sounds are decomposed by the auditory system (Cusack & Carlyon, 2003). Throughout the history of auditory research, advancements in the understanding of hearing have arisen from the close interaction between psychophysics and electrophysiology. Response times are a common behavioral measure of mental representation and are often used as an indicator of processing demands and computational load (e.g. Donders, 1868; Cattell, 1886; Luce, 1986; Pouliot & Grondin 2005). Here I will argue that in some cases, subjects' behavioral performance alone might lead to erroneous conclusions about the nature of the processing involved.

The major difficulty in trying to understand the computations that underlie the organization of sensory information is that perception is fast: perceivers are usually unaware of the stages of processing by which sensory information is converted into an internal image of the world. Arguably, psychophysics, as it requires subjects' conscious participation, only taps the final representational stages. To study how perceptual representations are created from sensory input, I employ a combination of

73

psychophysics and brain imaging (Magnetoencephalography; MEG). MEG, the magnetic counterpart of the more widely used EEG method, is a non-invasive neurophysiological technique that measures the magnetic fields generated by the neuronal activity of the cerebral cortex (Lu & Kaufman, 2003). Unlike fMRI or PET which are indirect measures of brain function reflecting brain metabolism, MEG is a direct measure of cortical activation and has a very high temporal resolution (events on the order of milliseconds can be resolved). Time-locked stimulus evoked cortical responses measured with MEG are characterized by increases in activation (peaks) at different latencies. Changes in this activation pattern (as a result of modifying the stimulus or the listener's attentional state) indicate changes within the neural mechanisms that these peaks reflect. Variations in latency can be interpreted as a slowing down or speeding up of a particular process, likewise changes in amplitude may reflect a difference in processing demands (e.g. Friederici, 2005). At present, because the mechanisms that sub-serve the construction of auditory perceptual representations are not well understood, our hypotheses are mostly related to time (e.g 'in the process of figure-ground segregation, computation A has to occur before computation B'). With its fine temporal resolution, MEG is especially suited for addressing questions related to the timing of brain activation. Reaction-time measurements are likewise well suited to measure the time course of behavioral responses. By studying how the dynamics of brain responses are related to behavior, I explore how behaviorally relevant representations arise from sensory input. This technique is particularly useful for dissociating bottom-up stimulus-driven processes from those that are affected by the perceptual state of the listener. In the visual

system, these kinds of questions have been the focus of experimental research for quite a long time (e.g Vecera et al. 2004; Peterson, 1999; Robertson, 2003; Amano et al., 2005), but they are just beginning to attract attention in the field of auditory neuroscience (Kubovy & van Valkenburg, 2001; Carlyon, 2004; Griffiths & Warren, 2004; Nelken, 2004; Scott, 2005).

 Here, I present the results of a study where the simultaneous acquisition of MEG and behavioral data revealed a curious incongruence between behavior and electrophysiology. The dissociation between subjects' perception, as reflected by their behavior, from lower level brain responses reveals implicit (pre-attentive) brain function related to the construction of auditory perceptual representations.

The specific processes that will be discussed here are related to the extraction of objects from background noise. The division of the scene into a figure (onto which attention is focused) and background is an important concept in scene analysis (Carlyon, 2004). These are the processes that underlie listener's ability to follow a friend's voice in a noisy party or to detect (and hopefully avoid) an approaching car in a busy street. In order to study the properties of these neural mechanisms I use simpler signals - faint tonal objects in loud background noise - that are intended to model more complicated real-life situations. By measuring subjects' brain activity while performing a tone detection task, I attempt to elucidate the mechanisms that enable listeners to recognize the tonal signals amidst the background.

In the experiments described below, we recorded brain activity while subjects listened to 1500ms long wide-band noise stimuli. In half of the presentations, a faint 500ms tone, with variable frequency, appeared at 1000ms post onset (Figure 2.10).
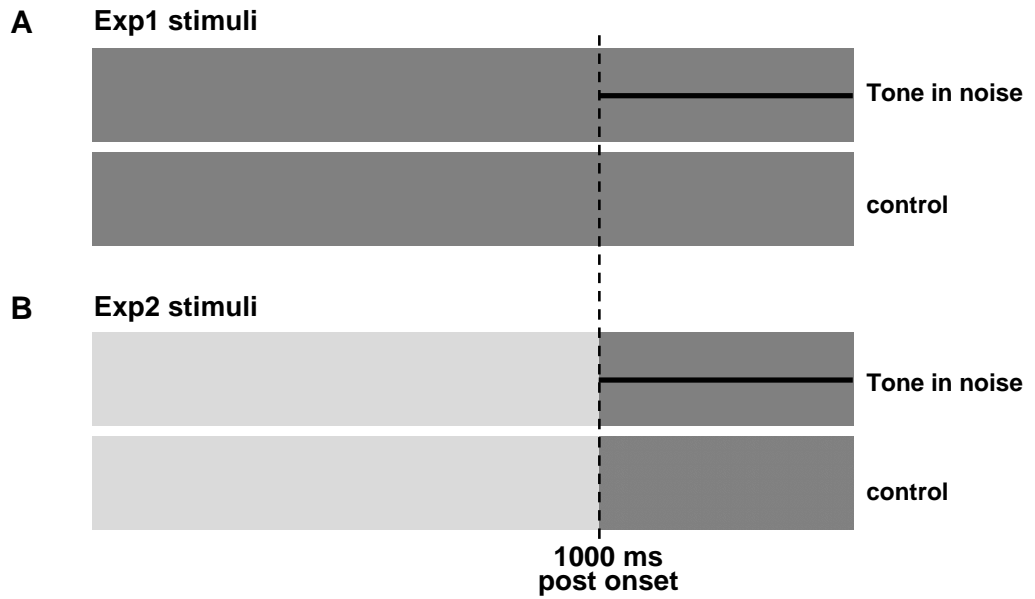
Figure 2.10. Schema of the stimuli used in the two experiments. (A) Stimuli for Exp1 consisted of 1500ms correlated wide-band noise (dark grey) with a 500ms faint tonal object (TN; black line) appearing at 1000ms post onset. Control stimuli were 1500ms long wide-band correlated noise. (B) Stimuli for Exp2 consisted of 1000ms of uncorrelated wide-band noise (light grey) followed by a 500ms long correlated noise segment which either contained a tonal object (target condition) or did not (control condition). Crucially, the last 500ms of the stimuli of Exp1 and Exp2 were identical

Subjects were instructed to press a button as fast as they can when they hear a tone popping out of the noise. The data reported here are a subset of the experimental data previously reported in (Chait et al., 2006). Those experiments were designed to investigate the process by which information from the two ears is combined to create a single coherent percept. We compared the cortical and behavioral responses to Huggins Pitch (HP; Cramer & Huggins, 1958), a stimulus requiring binaural processing to elicit a pitch percept, with responses to tones embedded in noise (TN)—perceptually similar but physically very different signals. Here we focus only on a portion of the signals studied in the original experiments.

In Experiment 1 (Figure 2.10A) the background noise was interaurally correlated (same noise signals presented to the two ears). The stimuli sound like a loud noise source (perceptually somewhat like radio static) in the center of the head with a faint tone appearing at 1000ms post onset. In order to investigate the effect of changing background on tone extraction, the stimuli of Experiment 2 (Figure 2.10B) consisted of interaurally uncorrelated noise (different noise signals presented to the two ears) which changed into correlated noise at 1000ms post onset. These signals sound like a diffuse noise source that changes into a centered noise source. Half of the time, a faint tonal object appeared at the same time as the noise-change. The simultaneous recording of behavioral reaction times and MEG response latencies to these sounds enables the investigation of the dynamics of the process by which the physical signals are separated into a perceptual representation consisting of a figure (the tone) and a background (the noise) and the degree of correspondence between behavioral and electrophysiological measures.

2.2.1 Methods
Subjects

Twenty subjects (mean age 24.6 years), took part in Experiment 1. Sixteen subjects (mean age 23.8) took part in Experiment 2. Twelve listeners participated in both experiments. All subjects were right handed (Oldfield, 1971), reported normal hearing, and no history of neurological disorder. The experimental procedures were approved by the University of Maryland institutional review board and written informed consent was obtained from each participant. Subjects were paid for their participation.

Stimuli

The stimuli in Experiment 1 were 1500 ms long: 1000 ms of interaurally correlated white noise (diotic stimuli) followed by a 500 ms tonal-object (200, 400, 600 or 1000 Hz) embedded in noise or by 500ms of the same correlated noise (control condition). The tonal objects in noise were either: a pure tone (TN), or a perceptually similar dichotic-pitch signal (HP). As stated above, the data reported here will concern only the pure tones (TN stimuli). Full results are described in (Chait et al., 2006). The stimuli of Experiment 2 were identical to those of Experiment 1 except that the initial noise was interaurally uncorrelated (statistically independent signals in each ear). In total each subject heard 100 presentations of each of the eight pitch conditions (HP 200, 400, 600, 1000 Hz; TN 200, 400, 600, 1000 Hz) and 800 (50% of all) presentations of the control stimulus. The order of presentations was randomized, with the inter-stimulus interval (ISI) semi-randomized between 500-2000 ms. The noises were created off-line by choosing 16000 pseudo-random (16 bit) numbers from a Gaussian normal distribution (sampling frequency 16 kHz), giving a

broadband signal up to 8000 Hz. To reduce the chance that observed effects are due to a specific instance of the white noise signal, 4 different instantiations were used. The stimuli were gated on and off using 15 ms cosine-squared ramps, with no gating at 1000ms post onset.

Procedure

The recording (~ 1.5 hours) consisted of two parts. First, subjects heard 200 repetitions of a 1 kHz (50 ms) sinusoidal tone (inter-stimulus interval randomized between 750-1550 ms) as a pre-experiment. Then subjects listened to the noise stimuli and performed a pitch detection task (50% of trials in Experiment 1 and 2) by pressing a button held in the right hand, as soon as they heard a faint tone pop out of the noise (tones appeared at 1000 ms post onset). Subjects were instructed to respond as quickly and as accurately as possible. The stimuli were presented in 10 blocks (approximately 8 minutes long) of 160 random stimuli.

Neuromagnetic recording and analysis

The magnetic signals were recorded with a 160-channel, whole-head axial gradiometer system (KIT, Kanazawa, Japan). The pre-experiment data (1 kHz sampling rate) were filtered from 1 Hz and 58.8 Hz, baseline corrected to the 100 ms pre-onset interval and stored in 500 ms (100 ms pre-onset) stimulus-related epochs. The data for Experiments 1 and 2 (1 kHz sampling rate) were filtered between 1 and 200 Hz with a notch at 60 Hz, and stored for later analysis. Raw data were noise-reduced using the Continuously Adjusted Least-Squares Method (Adachi et al., 2001).

In the pre-experiment, auditory evoked responses to the onset of the pure tones were examined and the M100 response identified. The M100 is a prominent and robust (across listeners and stimuli) deflection at about 100 ms post onset, and has been the most investigated auditory MEG response (see Roberts et al., 2000 for review). It was identified for each subject as a dipole-like pattern (i.e. a source/sink pair) in the magnetic field contour plots distributed over the temporal region of each hemisphere. In previous studies, under the same conditions, the resulting M100 current source localized to the upper banks of the superior temporal gyrus in both hemispheres (Lütkenhöner & Steinsträter, 1998). For each subject, the 20 strongest channels at the peak of the M100 (5 in each sink and source, yielding 10 in each hemisphere) were considered to best reflect activity in the auditory cortex and thus chosen for the analysis of the experimental data.

In Experiments 1 and 2, 1500ms long epochs (50ms pre onset) were extracted for each condition Epochs with amplitudes larger than 3 pT were considered artifactual and discarded from further analysis, resulting in the rejection of ~ 10% of the trials. The rest were averaged, low-pass filtered at 20 Hz and base-line corrected to the full range of the epoch. In each hemisphere, the root mean square (RMS) of the field strength across the 10 channels, selected in the pre-experiment, was calculated for each sample point.

2.2.2 Results

MEG data

The stimuli of Experiment 1 (Fig. 2.10A) sound like a 1500 ms continuous noise located in the center of the head. In half of the trials, a faint tonal object appears amidst the noise at 1000 ms post onset. The initial portion of the stimuli of Experiment 2 sounds like a diffused noise (the binaural stimuli are not fused to a unitary auditory object). At 1000 ms the noise changes from diffused to centered, and at the same time, in half of the trails, a faint tonal object appears (Fig. 2.10B). Here we focus on the brain activity occurring after the appearance of the tonal object, at 1000 ms post stimulus onset.

Experiment 1

In Experiment 1, we identified a cortical response, at approximately 160-200ms post tonal onset, evoked by the appearance of a tone in the noise. Figure 2.11 shows this response for the 200 Hz tone as an example. That this response is indeed related to the appearance of the tone is evident from the fact that it does not show in the control (no tone) condition and that it is affected by the frequency of the tonal object.[4] In Chait et al., (2006) we hypothesize that this increase in current is generated by the neural mechanisms that are responsible for the extraction of tonal objects from background noise. Alain et al. (2002; see also Dyson & Alain, 2004), in EEG studies of concurrent sound segregation, reported that the perception of a mistuned harmonic as a separate sound is associated with a negative wave peaking at about 150 ms after

---

[4] Tones of different frequencies result in a response with a slightly different latency, such that low frequency tones evoke later responses than high frequency ones. See Figure 6 here and further discussion in Chait et al. (2006).
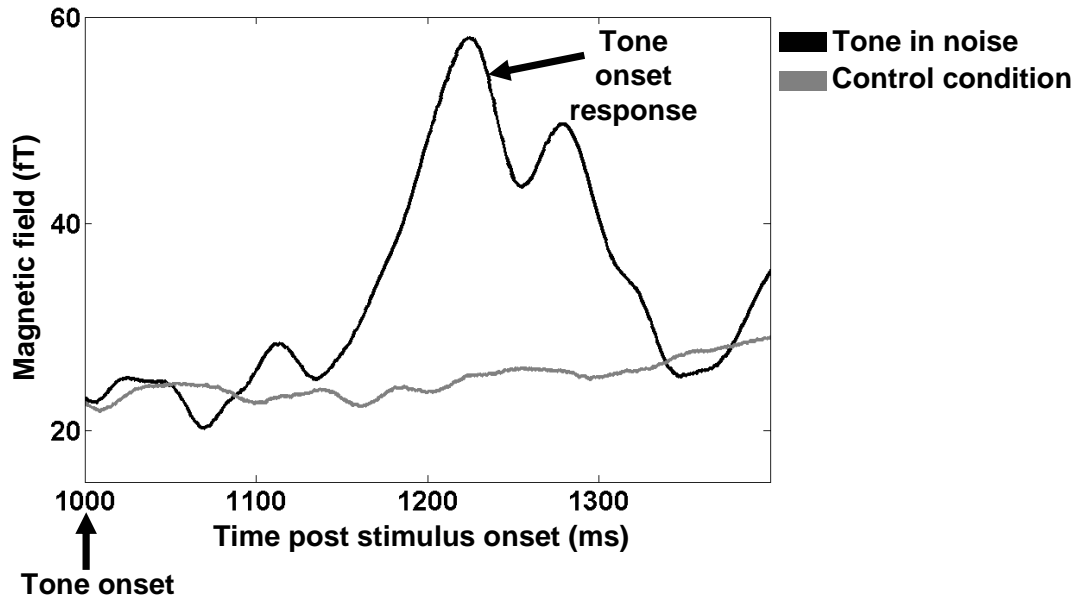
Figure 1.11: Electrophysiological data from Experiment 1. The group-RMS (RMS of individual subject RMSs) in the left hemisphere for the 200Hz tone condition, in black (all other conditions showed similar responses) and the control condition, in grey. The cortical response to the appearance of a tone amidst the noise was characterized by a magnetic deflection at about 160-200ms post tone onset. No such deflection appears in the control condition. To facilitate comparison between the two conditions, and to compensate for the fact that the control condition was averaged over many more (800 vs. 100) repetitions, the average amplitude (DC) of the control condition was adjusted (by 7 fT) to match that of the tone condition
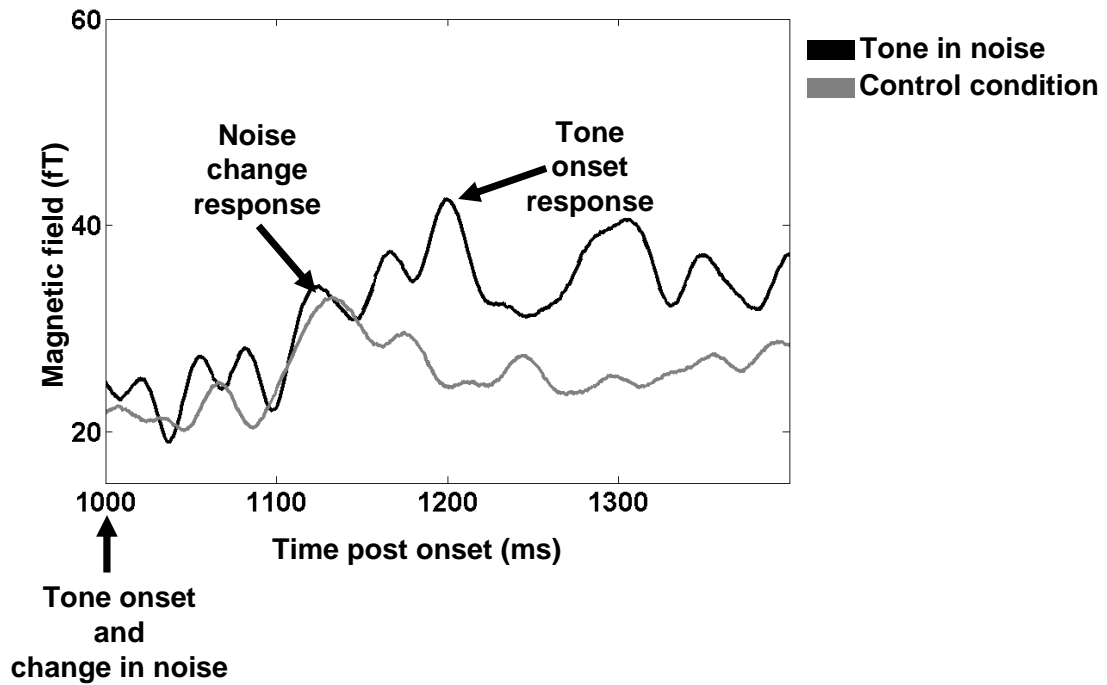
Figure 2.12: electrophysiological data from Experiment 2. The group-RMS (RMS of individual subject RMSs) in the left hemisphere for the 200Hz tone condition, in black (all other conditions showed similar responses) and the control condition, in grey. The response to the change in the noise (at 1000ms) is evident at ~1130ms in the control condition. In the tone condition, the response to the change in the noise and the response to the change in the tone are distinguishable. Thus, even though the appearance of the tone and change in the noise were simultaneous in the signal, these were processed at different times in cortex. To facilitate comparison between the two conditions, and to compensate for the fact that the control condition was averaged over many more (800 vs. 100) repetitions, the average amplitude (DC) of the control condition was adjusted (by 7 fT) to match that of the tone condition

sound onset. In their stimuli, the onset of mistuning coincides with that of the stimulus, so the response components of the two cannot be isolated but the properties of the wave (referred to as object-related-negativity) are very similar to those observed in the present study.

Experiment 2

The results of Experiment 2, are plotted in Figure 2.12 (as in Figure 2.11, we plot the 200Hz tone data as an example). The control stimuli of experiment 2 (Fig 2.10B) contained a change in the noise at 1000ms post onset. The neural response to this change is reflected in a peak at 1130ms (130ms post change), which is visible in the control condition data (grey line in Fig 2.12). As discussed above, the tonal conditions contained two simultaneous changes: a change in the interaural correlation of the noise, and the appearance of the tone. Remarkably, as can be seen from Figure 3, the responses to these changes happen at different times, and we observe two separate peaks that correspond to the change in the noise, at 1130ms, and to the onset of the tone, at 1200ms, respectively.

Our ability to segregate the neural responses stems from their temporal separation (response to the onset of the tone is about 50-100ms later than the response to the change in the noise, depending on the specific frequency of the tone), the fact that the noise change peak is not modulated by the tonal frequency, and the fact that the two responses have different scalp distributions (the response to the onset of the tone is strongly left lateralized, whereas the response to the change in the noise is

right lateralized), see Chait et al. (2006) for more details. The finding that the change in the noise and the appearance of the tone, though happening simultaneously in the signal, are processed separately, by different mechanisms (as indicated by the different scalp distributions of the responses) is consistent with Nelken (2004), who hypothesizes that in auditory cortex, sound is not represented purely in terms of it's physical features but rather in terms of auditory 'objects' – mental representations of the auditory sources that compose the input to the ear.

Comparing behavioral and MEG data

Although the 'tone onset response' is related to the extraction of the tonal target, we find no correlation between individual subjects' behavioral response times and MEG latencies, in either experiment. This is consistent with previous reports that the latency of early ERP components such as N100 and P200 are not correlated with response times (see also chapter 3 here). Several studies (e.g. Kutas et al., 1977) suggest that brain activation starting from around 300ms post onset may begin to be correlated with behavioral responses.

An interesting pattern of results is revealed when examining how behavioral and MEG response latencies change between Experiments 1 and 2. Behaviorally, performance in Experiment 2 was worse than in Experiment 1. Those subjects that participated in both experiments (N=12) observed that Experiment 2 was more difficult. They reported that in addition to the change in the noise that occurred simultaneously with the appearance of pitch and hindered the detection, the quality of
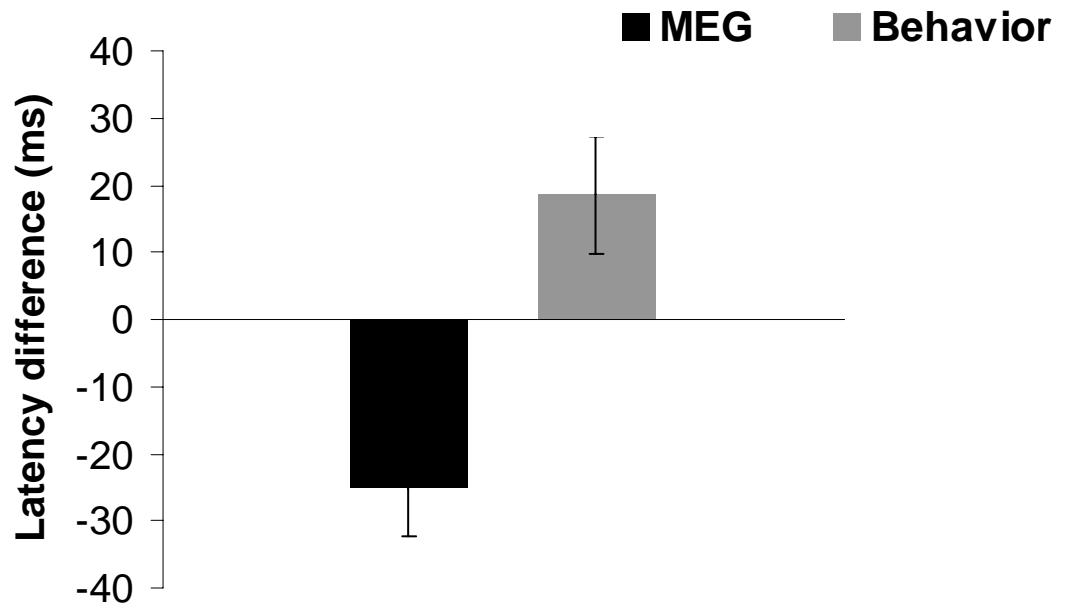
Figure 2.13: Average response time differences (collapsed over frequencies) between Experiment 2 and Experiment 1 for the 12 subjects common to both experiments. Positive values indicate responses in Experiment 2 that were delayed relative to Experiment 1. Whereas behavioral responses (grey bar) were longer in Experiment 2 than in Experiment 1, electrophysiological responses to the same stimuli showed an opposite pattern and were earlier in Experiment 2 than Experiment 1

**~450 ms
post onset**

behavior

**Higher level
decision
mechanisms**

**160-200 ms
post onset**

object 1

Tone

object 2

Noise

Figure 2.14. Schema of the hypothetical relationship between behavioral responses measured with psychophysics, and brain responses measured with MEG. The MEG responses measured at 160-200ms post onset reflect a representation in which the information about the tone and about the noise are already segregated and processed by different mechanisms (bottom squares). The outputs of these computations are combined at a later stage that underlies subject's conscious decision about the existence of the tone (top square). The uncorrelated leading noise in Experiment 2 causes a faster response (relative to Experiment 1) in the lower level 'tone detecting; stage (see Chait et al (2006) for details) and a slowing down (relative to Experiment 1) at the higher level decision mechanisms. This slowing down is a result of the additional computational load, caused by the task-irrelevant information about the change in the noise

the noise (as two unfused objects at each ear) in the first 1000 ms of the stimulus made it harder to detect the tonal object. This difficulty was reflected in increased response times in experiment 2 relative to experiment 1 ($p<0.028$). On average, response times were about 20ms longer (see Figure 2.13). However, when comparing the latency of the brain responses to the onset of the tonal objects,[5] we observe the opposite pattern: The latency of the tone onset peak in Experiment 2 is earlier than in Experiment 1 for all tonal frequencies used ($p<0.01$). Thus behavioral and electrophysiological measures showed a fundamental incongruence (Figure 2.14): listeners *unconsciously* detected the tones in Experiment 2 faster than in Experiment 1, but *consciously* reported difficulties and detected the tones in Experiment 2 slower than in Experiment 1.

Implications for processing

The fact that the behavioral difficulty in Experiment 2 was not apparent in the brain responses we measured at ~150ms post tone onset suggests tha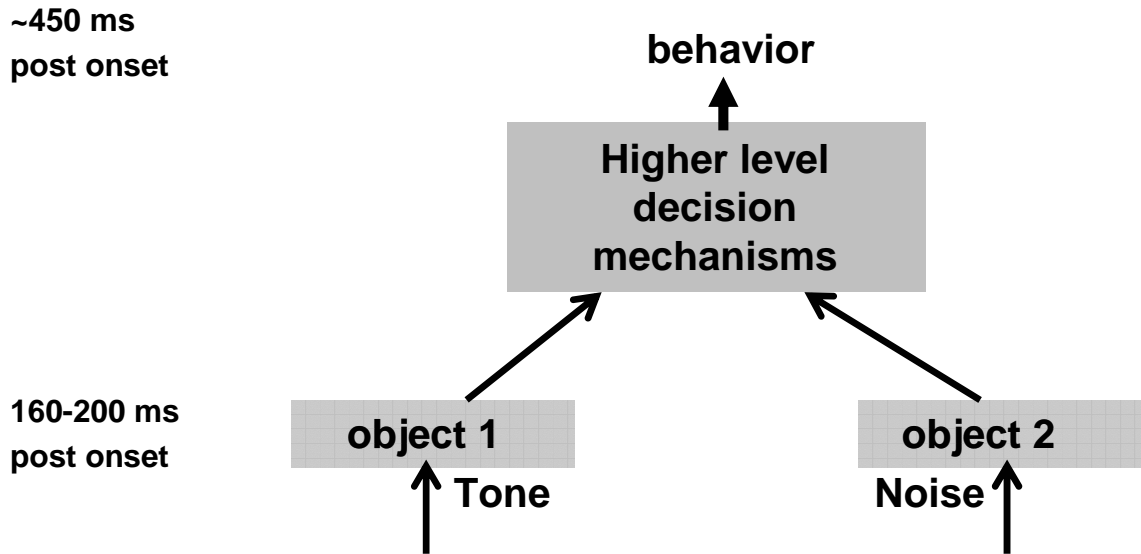t behavior is affected by later processing stages. Figure 2.14 presents a schema of the hypothetical relationship between behavioral responses measured with psychophysics, and brain responses measured with MEG. In the lower level scene analysis stages, which correspond to the cortical responses we recorded, the change in the noise and the appearance of the tone are processed separately (bottom squares). These mechanisms receive information from earlier processing stages where the uncorrelated noise (in

---

[5] Electrophysiological response latencies are defined as the time at the peak of the response. For example, the latency of the tone onset response in Figure 1 is 1220 ms (220 ms post tone onset), whereas the latency of the tone onset response in Figure 2 is about 1200 ms (200 ms post tone onset).

Experiment 2) aids the detection of the tonal objects.[6] The information from the tone and noise change-detectors is combined in a higher level, 'conscious' stage (top square) that makes the decision regarding whether a tone has actually appeared. This is the level that directly affects the initiation of behavior. A concurrent (irrelevant) change in the noise introduces an extra computational load at this stage and thus a slowing down is observed. This is in contrast to the neural detection task itself, which is not slowed down (and is in fact faster).

Figure 2.15 presents further evidence for the indirect relationship between behavior and the MEG deflections we measured. When comparing the trends of the MEG responses in Experiment 1 and 2 (black solid and dashed lines) we find that they are not correlated: response latency difference varies with frequency.[7] In the case of a linear relationship between behavior and the lower-level mechanisms reflected in the MEG responses, we would expect behavioral data to exhibit a similar pattern. However, behavioral responses in Experiment 1 and 2 (grey solid and dashed lines) are significantly correlated (Pearson's correlation coefficient $\rho=0.961$ $p=0.039$), implying a similar latency shift for all frequencies.

Similar effects are observed when computing the correlation of response latencies across experiments for each frequency separately (Figure 2.16). In the case of MEG responses (Figure 2.16A), we find that none showed a linear interaction

---

[6] The change in the interaural configuration of the leading noise caused an *earlier* brain response to the appearance of the tone in Experiment 2 relative to Experiment 1. This response latency difference might originate as early as the Superior Olivery Complex (SOC) in the brain stem, where binaural processing begins. This is discussed in detail in Chait et al (2006).
[7] This is possibly because the differing physical aspects of the stimuli in the two experiments differentially affected the processes underlying the tone extraction mechanisms.
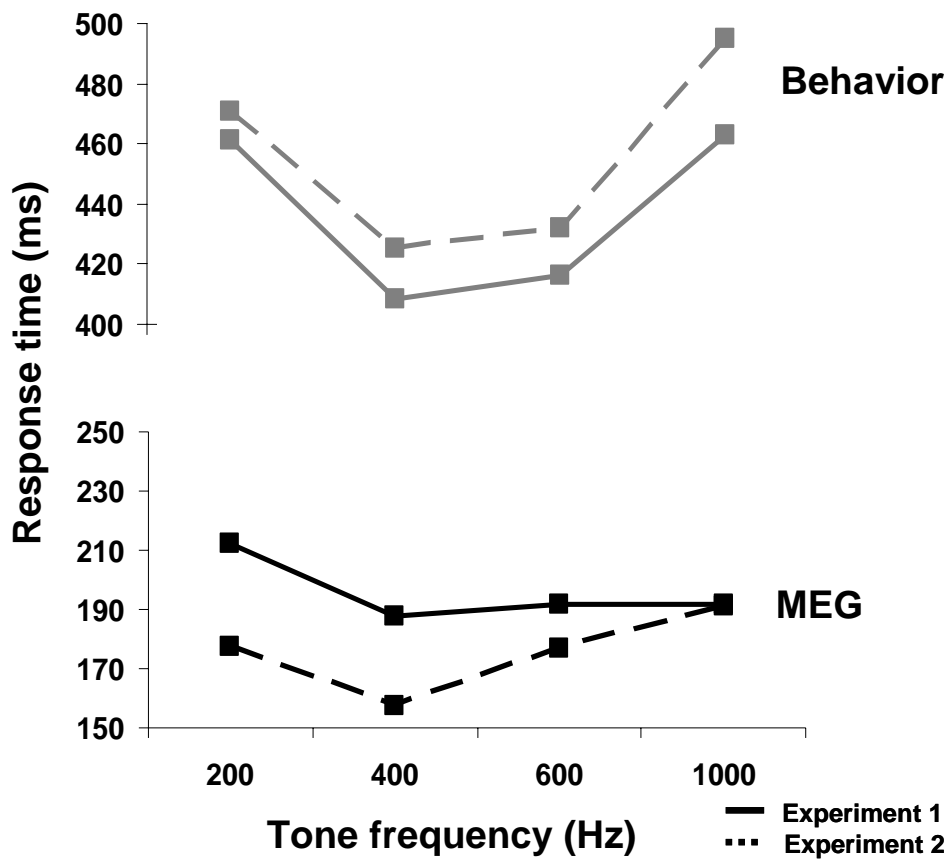
Figure 2.15: A comparison of behavioral response times (top) and MEG response latencies (bottom) for the 12 subjects common to both experiments. Experiment 1 data: solid lines; Experiment 2 data: dashed lines. MEG and behavioral responses in either experiment are not correlated. So is the relationship between the brain data of Experiment 1 and 2 (black solid and dashed lines). However, there is a strong correlation between behavioral responses in the two experiments (grey solid and dashed lines are parallel). Also visible in this figure are the increased behavioral response times and decreased MEG response latencies between Experiments 1 and 2
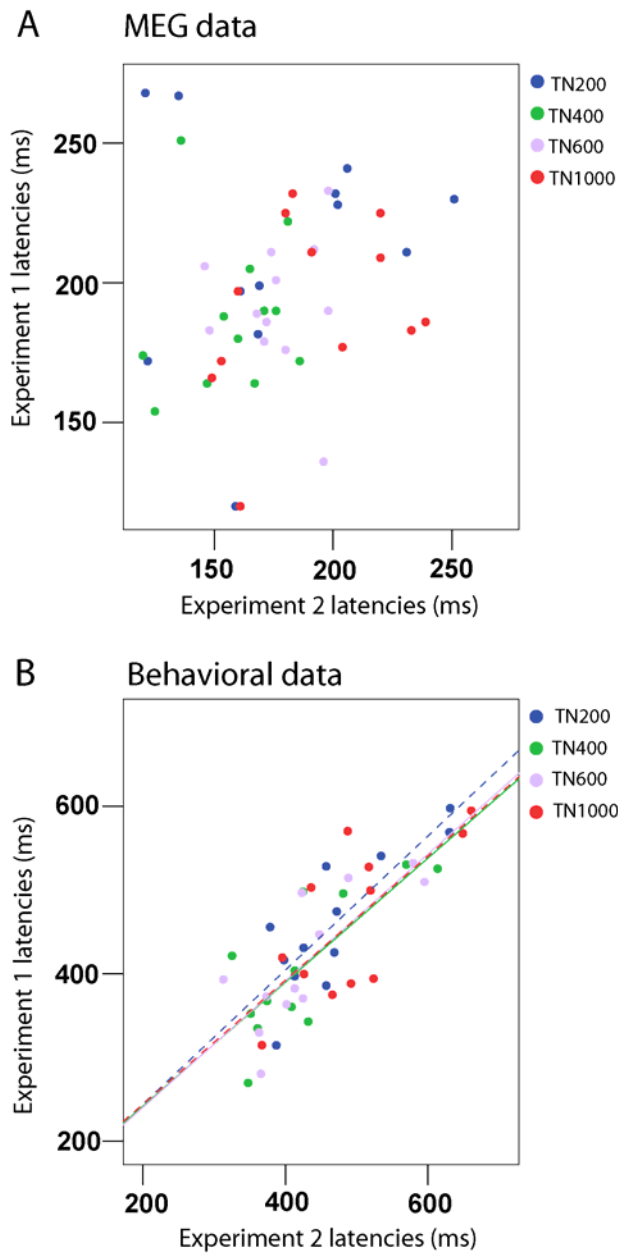
Figure 2.16: Scatter plot of the individual MEG latencies (A) and behavioral response times (B) in Experiments 1 and 2 for each of the tonal-object frequencies. Whereas MEG response latencies show no correlation, Behavioral response times between the two experiments are highly correlated. Dashed lines in (B) are linear fits (all with an R Sq value of about 0.6).

between response latencies in Experiment 1 and Experiment 2. However, the same analysis on the behavioral data (Figure 2.16B) reveals a strong linear relationship between the response times in Experiment 1 and Experiment 2 for all frequency conditions (200 Hz: $\rho=0.826$ $p=0.001$, 400 Hz: $\rho=0.775$ $p=0.003$; 600 Hz: $\rho=0.789$ $p=0.002$; 1000Hz: $\rho=0.723$ $p=0.008$). Such a consistently strong correlation is quite surprising in light of the fact that the two experiments took place at least a month apart. At the same time, it is known that individual response times tend to be stable over time (e.g. Cattell, 1886).

These data thus suggest that the difference between the stimuli of the two experiments had a non-linear effect on cortical response latencies, but a linear effect on behavioral response times. This pattern is consistent with the mechanism described in Figure 2.14: the leading uncorrelated noise in Experiment 2 has effects on different levels of processing: (1) It affects low level (sub-cortical) mechanisms related to extracting the tone from the noise (see Chait et al. (2006)), and so may conceivably affect different frequencies in a different way. These mechanisms probably relay information to the cortical responses we observe ('object 1' block in Figure 2.14). (2) The change in the noise also activates the cortical noise-change detectors ('object 2' block in Figure 2.14). Even though the change in the noise is irrelevant to the subjects' task, we show that it influences their performance. The effect of the irrelevant noise change on the higher level decision mechanisms is a general increase in processing load (the information that the noise has changed is identically salient in all conditions) which results in a constant latency difference between Experiment 1 and Experiment 2.

2.2.3 Discussion

In the case of the experiments presented here, behavior and electrophysiology, if studied separately, might lead to different conclusions about the nature of the processing involved. The simultaneous acquisition of both MEG and behavioral data puts stronger constraints on the interpretation, revealing a multi-staged process of the construction of perceptual representations. These findings contribute to our understanding of auditory scene analysis in several ways: They demonstrate that at approximately 150 ms post onset the change in the tone (figure) and the change in the noise (ground) are already processed separately. Additionally, they reveal at least two stages in the detection of auditory objects, and provide a time frame in which each of the stages operates. Behavioral response times in this study were about 450ms post tonal onset. Since it takes about 150ms to program a motor command (e.g. Sereno &Rayner, 2003), we are faced with a 'missing 100 milliseconds' (from the peak of the cortical responses to when button-press motor commands are initiated). In order to study the nature of the computations that are executed between the subconscious detection of the tone and the conscious decision, we can manipulate the subject's task or the early cortical responses (by changing the properties of our stimuli) and examine the effect on reaction time. For example, a possible prediction is that even in the case of three concurrent changes in the stimulus, the low level stages should remain unaffected but behavioral responses should exhibit an even grater delay (more interference).

That there are distinct stages of processing between target detection and behavioral response is not, in itself, a surprising finding (although it would have

probably surprised Cattell (1886) who believed that simple reaction times, such as those measured here do not involve the cortex at all)[8]. MEG and EEG studies are consistently observing that early brain response latencies are not correlated with response times. This is related to the fact that the latency of early brain responses is less variable (between and within subjects) than behavioral response times (Thompson et al., 1996; Sato et al., 2001; Carpenter, 2004). What is noteworthy in the current results is the degree to which behavioral data and electrophysiological data are incongruent.

The observations here are consistent with a series of studies on the properties of saccadic latencies. An interesting property of behavioral response times (in general, and saccades in particular) is that they tend to be distributed randomly (Luce, 1986; Carpenter, 1999a). This effect is attributed to the way decision mechanisms operate, that is, to the processes underlying a listener's judgment that a target has indeed appeared (Carpenter, 1999b). According to a model suggested by Carpenter and colleagues (Carpenter 1999a,b; Carpenter, 2004), information arriving from the systems responsible for target extraction causes a decision signal to rise linearly from an initial level to a threshold level, at which the response is triggered. This is the processing stage at which a subject supposedly becomes 'consciously' aware of the stimulus. What causes the variation in response times is that the rate of rise on different trials varies with a Gaussian distribution. Fairly convincing evidence for

---

[8] Cattell (1886) summarizes his simple reaction time experiments by saying that: "From these experiments we see that ordinary degrees of attention do no greatly affect the length of the reaction time. We find, further, grounds for assuming that the cortex is not concerned and that perception and willing are not factors of the reaction-time. It is not necessary to perceive the stimulus before the motor center can be excited; and the willing – not of necessity given in consciousness- is done before the stimulus occurs, and consists in setting the brain-parts concerned in a state of readiness", pp 242.

such a stochastic mechanism has been obtained from awake monkey electro-physiology (Hanes & Schall, 1996; Thompson et al., 1996) and from saccadic decision tasks in humans (Reddi & Carpenter, 2000; Reddi et al., 2003). This body of research also provides evidence for the independence of detection and decision (Carpenter, 2004), as suggested by the results here as well.

In order to demonstrate the existence of two mechanisms of this kind, that precede the initiation of behavior, studies have independently varied detection (by manipulating the salience of the target) and decision (by manipulating parameters related to prior probability or the way responses are to be executed[9]). The different conditions in the present series of experiments can also be interpreted in this way. Here, the uncorrelated noise in Experiment 2 serves both to *facilitate* the extraction of the target (the tone) and to *impede* the decision about its appearance. Importantly, our results imply that the dissociation between detection and decision can be achieved under more natural conditions (without the need to manipulate response interference), where in addition to the target another, task-irrelevant, object competes for attention. The interfering object, depending on its properties, might or might not affect target salience, but will affect the decision mechanisms.

These findings are interesting from an auditory processing point of view because they reveal previously unobserved stages in the re-construction of the auditory scene from sensory information, and provide a methodological tool for a further, more rigorous, examination of how the different stages interact to affect behavior. These results may also be of interest to researchers of higher level human

---

[9] For example, asking subjects to press the right button when they detect the word 'RIGHT' and the left button when they detect the word 'LEFT' (easy task) or vice versa (harder task) (McCarthy & Donchin, 1981).

faculties, such as decision making and reasoning (e.g Gold & Shadlen, 2001), because they reveal a method by which one can measure, in humans, the (pre-attentive) inputs into the decision system and how they affect its output. For instance, by manipulating the salience of auditory cues and parametrically affecting the outputs of the lower level cortical mechanisms (bottom squares in Figure 2.14) while at the same time measuring behavioral reaction times, one may be able to infer the kinds of reasoning rules and heuristics that subjects are employing in the process of making a response.

2.3 Dyslexic listeners exhibit delayed detection of tonal targets in background noise: implications for processing[10]

Developmental dyslexia is a genetically based neurological disorder (Grigorenko et al., 2001; Wood & Grigorenko, 2001; Meng et al., 2005; Ramus, 2004) characterized by impaired reading despite adequate intelligence and educational opportunity. These difficulties in acquiring literacy have been associated with a more fundamental deficit in phonological processing (the internal representation of speech sounds) and phonological awareness, such as the ability to isolate and manipulate the constituent sounds of words (Bradley & Bryant, 1983; Snowling et al., 1986; Torgesen et al., 1994; Snowling, 2000; National Reading Panel, 2000). In addition to a wide variety of phonological deficits, individuals with dyslexia are often (but not always) impaired in a range of non-linguistic auditory and visual processing tasks (Eden et al., 1996; Witton et al., 1998; Wright et al., 2000; Stein, 2001; Hari & Renvall, 2001; see also Ramus 2003). The relationship between these lower level sensory deficits and reading impairment is unclear and extensively debated (Hari & Renvall, 2001; Ramus, 2003; Ramus 2004).

Among the wide range of auditory deficits linked with dyslexia (for reviews see Wright et al., 2000; Rosen, 2003; Ramus 2003), several studies have reported deficits associated with low level binaural processing and specifically with binaural

---

[10] This section is based on Chait M, Eden G, Poeppel D, Simon JZ, Hill D, Flowers L (submitted) Dyslexic listeners exhibit delayed detection of tonal targets in background noise: implications for processing.

unmasking. Binaural unmasking (Grantham, 1995), refers to the fact that a signal masked by correlated noise (identical noise signals at the two ears) can be made easier to detect by inverting the signal, or the noise, in one ear. That is, when the phase of the signal at the two ears is not the same as the masker. In natural environments, sound sources at different spatial locations result in different interaural phase configurations and binaural unmasking is thought to be one of the mechanisms that facilitate listeners' ability to detect objects in noisy environments (e.g. being able to follow a conversation in a noisy party; Cherry, 1953). Binaural unmasking is based on the peripheral preservation of accurate timing information of the signal at each ear and on mechanisms that are sensitive to the similarity of the signals at the two ears. Cells exhibiting these properties are found in the Medial Superior Olive (MSO) of the brainstem. In animal studies, the Inferior Colliculus (IC), the projection target of the MSO, exhibits correlates of binaural unmasking (Jiang et al., 1997ab; Palmer et al., 2000). McAnally & Stein (1996) found reduced binaural unmasking in dyslexic listeners. This has been attributed to impaired low level binaural integration mechanisms and interpreted as a manifestation of a general dyslexic disorder related to processing of temporally fine sensory information (Stein, 2001). Similar findings of impaired binaural facilitation of signal-from-noise extraction in dyslexia have been reported in Dougherty et al (1998), using dichotic pitch stimuli. However, as is frequent in dyslexia research, other investigations (Hill et al., 1999; Amitay et al, 2002) reported no difference between control and dyslexic listeners' binaural unmasking thresholds.

The present study seeks to clarify this issue by assessing dyslexics' binaural processing abilities with a new behavioral paradigm. We compare the performance of dyslexic listeners and their matched controls in a tone detection task. The stimuli (figures 2.17, 2.18) consist of a tonal object appearing amidst wide-band background noise. The tonal object is either a pure tone (TN) or a Huggins Pitch (HP)—perceptually similar but physically very different signals. Huggins Pitch (Cramer & Huggins, 1958) is the auditory counterpart of the well known 'random dot stereogram' (RDS) effect in vision (Julesz, 1971). An RDS is a binocularly presented pair of images with randomly distributed dots. Each image appears random when viewed individually but when displayed simultaneously, one to each eye, they evoke a coherent 3-D percept. The illusion is created by presenting identical RDSs to the two eyes except that one image contains a group of dots that are slightly shifted relative to the other. The visual system fuses the shifted and non-shifted dots to create a 3D percept of an image (corresponding to the shifted dots) floating above the background (of the un-shifted dots). Similarly in audition, if a random broadband noise signal is presented to one ear, and the same random noise—but with a phase shift of π over a narrow frequency band—is presented to the other ear, this results in the perception of a faint tonal object with a pure tone quality (and a pitch that matches the center frequency of the phase-shifted band), embedded in noise. The input to either ear alone is just white noise, completely lacking any spectral or temporal cues to pitch. The fact that we are nevertheless able to perceive pitch when the two signals are presented dichotically implies that the HP percept is created by a mechanism that receives the inputs from
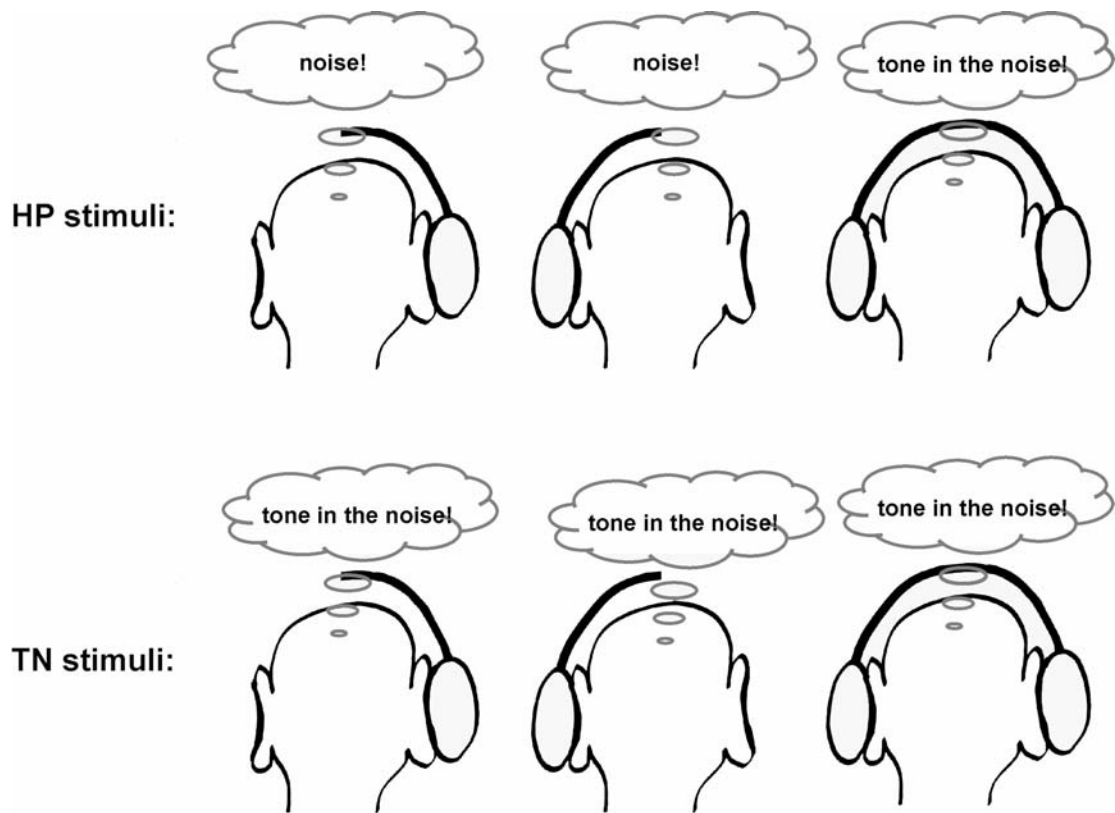
Figure 2.17: HP and TN stimuli evoke very similar percepts when listening with the two ears: a faint tonal object amidst background noise. The tone in the Huggins Pitch (HP) stimuli is only audible when listening with two ears. When listening with each ear alone, stimuli sound like white noise. The tone in the Tone in Noise stimuli is audible monaurally.
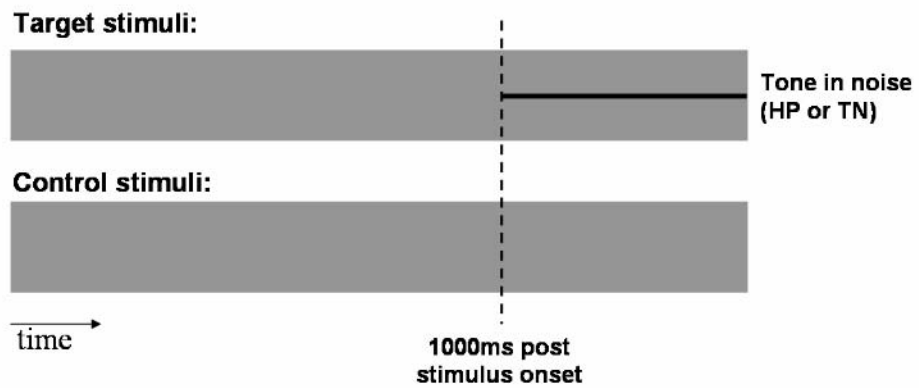
Figure 2.18: Schema of the stimuli used in the current experiment. Signals consisted of 1500ms correlated wide-band noise (dark grey) with a 500ms faint tonal object (TN or HP; black line) appearing at 1000ms post onset. Control stimuli were 1500ms long wide-band correlated noise.

the two ears, collates them, and emerges from the correspondence the percept of a tone (Figure 2.17).

Dougherty et al. (1998), using signals similar to HP, measured dyslexic listeners' perception of melodies consisting of dichotic pitch signals. Our Task is more straightforward: Subjects, who were not informed about the existence of different types (HP vs. TN) of tonal stimuli, performed a pitch detection task (50% of the trials) by pressing a button, held in the right hand, as soon as they heard a tone popping out of the noise (Figure 2.18). Reaction times (RT) and accuracy scores were stored and analyzed. This design is based on the fact that the pitch evoked by HP is perceptually similar to that evoked by a binaurally (diotically) presented tone in noise. A useful feature from the methodological point of view is that HP stimuli lack any structure (spectral or otherwise) that can be extracted at a low level within monaural pathways. Whereas the TN stimuli contain explicit pitch information that can be detected monaurally even at the level of the auditory nerve, the HP stimuli are indistinguishable from white noise up to the level of the MSO, where phase and timing information from the two ears are first compared.

We have previously used the same stimuli in a combined brain-imaging (with magnetoencephalography; MEG) and psychophysical investigation (Chait et al., 2006). In that study we identified an auditory cortical response, at approximately 160-200ms post tonal onset, evoked by the appearance of a tone (HP or TN) in the noise. These responses to HP and TN appeared to result from the same cortical processing mechanism but differed in their *timing*: (normal reading) subjects responded (behaviorally and electrophysiologically) approximately 30ms *faster* to HP

stimuli than to perceptually similar TN stimuli, regardless of frequency tested. This latency difference is attributed to low level binaural processes (as early as the MSO) that operate differently on the two kinds of signals (see Chait et al., 2006). These stimuli are therefore a useful tool to address the issue of dyslexics' alleged impaired binaural processing, and to differentiate between relatively high level impairments (for example, difficulties in extracting objects from noise) from low level deficits, related to combination of fine temporal information across ears. A low level deficit associated with dyslexic listeners should result in an altered pattern of latency differences between HP and TN stimuli, as compared to normal reading controls.

The current design carries several additional benefits: Most other dyslexia auditory processing studies use either multiple interval forced choice designs, requiring subjects to perform comparisons between stimuli (e.g Amitay et al., 2002; Hill et al., 1999; Ben-Yehudah et al., 2004; Wright et al., 1997) or designs that involve verbal labeling of the stimuli (e.g 'left'/'right' in Dougerhty et al, 1998). Our task is based on the simple detection of a signal. This method avoids alternative explanations related to impaired working memory in the dyslexic population (Marshall et al., 2001; Banai & Ahissar, 2004). Additionally, most investigations of auditory processing deficits in dyslexia are based on threshold measurement as a metric of performance (e.g Amitay et al., 2002; Hill et al., 1999; Ben-Yehudah et al., 2004; Wright et al., 1997). In the present study, we assess listeners' response times. Although response time measures and threshold measurements are correlated, they do not necessarily reveal identical information about processing (Luce, 1986).

Due to the variety of deficits exhibited by dyslexic individuals and the diagnostic criteria employed by different investigators, dyslexics are a notoriously heterogenic population. This variability often results in conflicting experimental findings. Here we study a relatively homogenous group of dyslexic adults, paired with controls from the same longitudinal cohort. Our results show no evidence of a low level, binaural processing impairment associated with dyslexia. We find, however, that the dyslexic subjects exhibited a general difficulty, manifested by a globally delayed response times, in extracting tonal objects from background noise.

2.3.1 Methods

Subjects

Seven dyslexic listeners (1 male; average age 24.4) and ten controls (3 male; average age 25.4) participated in the experiment. One control subject was excluded from analysis because of inability to perform the task (over 50% miss rate for both HP and TN targets). The subject profiles below refer to the 16 remaining participants.

Subject Selection Based on Childhood Testing and Adult Test Profiles

The participants were recruited from a longitudinal study of reading development that followed students from early elementary grades through high school. Mean ages of unaffected and affected readers at adult follow-up did not differ but there was a difference in education (15.1 and 13.3 years, respectively, *p*=0.02).

Affected status was defined by a non-word reading score below the 15th percentile (greater than one standard deviation below the mean) on a test of non-word

reading, either the Word Attack subtest of the Woodcock-Johnson Psycho-Educational Battery (WJ; Woodcock and Johnson, 1977) or the non-word reading portion of the Decoding Skills Test, part II (DST; Richardson and DiBenedetto, 1985). Where available, 3rd grade scores were used for these determinations, although for two dyslexic subjects it was necessary to use 1st grade scores. Note that cut scores for the WJ and DST were obtained from a subsample of 220 children, part of an original sample of 485 children who were recruited by stratified random sampling for ethnicity and socioeconomic status from a single city-county school system to study reading development. The smaller n=220 sample were those whose test batteries were complete across three testing cycles (at 1st, 3rd, and 8th grades) and preserved the ethnicity and ability levels of the original cohort. Reading and IQ scores were normally distributed. The adult follow-up test battery included measures in the following categories: Phonemic awareness was measured by the Test of Auditory Analysis Skill, adapted from Rosner, 1979, and expanded to 20 items (TAAS+), and the Lindamood Auditory Conceptualization test – 3rd Edition (LAC3; Lindamood and Lindamood, 2004). Non-word reading, was measured by the Woodcock Johnson III Word Attack subtest (WJ-III WA; Woodcock, McGrew, and Mather, 2001). Word and paragraph reading accuracy were measured by the Woodcock Johnson III Letter/Word Identification subtest (WJ-III WID; Woodcock, McGrew, and Mather, 2001) and the Gray Oral Reading Test, 3rd Edition, Accuracy score, (GORT3 Acc; Wiederholt and Bryant, 1992). Reading comprehension was measured by the Woodcock Johnson III Passage Comprehension subtest, (WJ-III PC; Woodcock, McGrew, and Mather, 2001) and the Gray Oral Reading Test- 3rd Edition,

Comprehension score (GORT-3 Comp; Wiederholt and Bryant, 1992). Fluency was measured by the Woodcock Johnson III Reading Fluency subtest, (WJ-III RF; Woodcock, McGrew, and Mather, 2001) and Rapid Automatized Naming (RAN; Denckla and Rudel, 1976). The WJ-III RF measures sentence reading fluency and the RAN tests fluent retrieval of highly familiar verbal labels for single letters, digits, colors, and objects. Separate RAN scores are calculated, to represent color and object naming (RANCO) and digit and letter naming (RANNL).

Also included was an instrument designed to briefly measure overall intellectual functioning. The Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) is a four-subtest measure consisting of two verbal and two nonverbal (performance) measures. The WASI generates a verbal intelligence quotient (VIQ) and a performance intelligence quotient (PIQ) as well as a composite, full-scale, IQ score (FSIQ), the latter being correlated with the Wechsler Adult Intelligence Scale-III FSIQ 0.92.

Dyslexic readers remained significantly poorer on adult follow-up testing on all tests of phonological processing, reading accuracy, and fluency (see table 1). Of particular interest is that they remained impaired on non-word reading and also on phonemic awareness (LAC3 and TAAS+), basic skills believed to be necessary for reading acquisition.

Stimuli

The stimuli (Figure 2.18) were identical to those used in Experiment 1 of Chait et al. (2006). The signals were 1500 ms long, consisting of 1000 ms interaurally

Table 2.1:
Comparison of typical and affected reader performance on adult cognitive measures.

Measure	Controls	Dyslexics	p-value
Demographics			
Education	15.11	13.29	0.019
Intellectual			
WASI VIQ	107.89	92.71	0.032
WASI PIQ	113.33	100.14	0.036
WASI FSIQ	111.67	96.14	0.009
Phonemic Awareness			
TAAS +	18.33	13.14	0.008
LAC3	104.22	81.00	0.003
Non-Word Decoding			
WJ-III WA	95.89	80.14	0.018
Word and Paragraph Reading Accuracy			
WJ-III WID	98.33	82.29	0.003
GORT-3 Acc	110.56	77.86	<.00001
Reading Comprehension			
WJ-III PC	99.89	96.29	0.441
GORT-3 Comp	103.33	98.57	0.666
Reading Fluency			
WJ-III RF	104.67	85.29	0.006
RANCO	109.56	85.14	0.004
RANNL	106.33	88.29	0.028

correlated white noise (identical noise signals at the two ears) followed by either HP or TN, as described below. Control stimuli consisted of a 1500 ms interaurally correlated noise.

We chose 4 center frequencies (200, 400, 600, 1000 Hz) that span the frequency region for which HP is salient. A 1000 ms sample of 'frozen' noise (sampling frequency 16 kHz, bandwidth 8 kHz) was generated for each of these 4 conditions. The HP signals were created by introducing a constant phase shift of π in a particular spectral region of the noise sample delivered to the right ear, while the original sample was delivered to the left ear (Yost et al., 1987). The width of the phase-shifted band was set to ±6% of its center frequency (Klein & Hartmann, 1981). The corresponding TN signals were produced by adding a pure tone (with one of the above frequencies) to the same noise samples used for the HP stimuli. Generally, listeners are able to match the HP signal to the pitch evoked by a pure tone (with a frequency that corresponds to the center of the phase-shifted band) with a standard error of about 3% (Klein & Hartmann, 1981). The amplitude of the pure tone signal was separately adjusted by 2 listeners to match the 'perceived tone' loudness of the corresponding HP stimulus.

When HP is perceived, the background noise is always localized to the center of the head but the tonal object may be reported as being at a lateral position away from the midline. It is lateralized to the left or to the right by some listeners but mostly evokes an inconsistent (ambiguous) lateralization, especially by inexperienced listeners (Yost et al., 1987; Zhang & Hartmann, 2004). Since no appreciable effects of perceived HP or TN lateralization were found in Chait et al. (2006), TN signals in

the current study were not matched to the perceived lateralization of HP and were always presented with the same amplitude of pure tone to both ears (perceived in the center of the head).

The stimuli were created off-line, gated on and off using 15 ms cosine-squared ramps (with no gating in the transition at 1000 ms post onset), and saved in 16-bit stereo WAV format at a sampling rate of 16 kHz. The signals were delivered to the subjects' ears with Sennheiser HD580 headphones and presented at approximately 75 dB SPL, to ensure a salient pitch. HP saliency increases with increasing noise level (Durlach, 1962).

In total each subject heard 50 presentations of each of the eight pitch conditions (HP 200, 400, 600, 1000 Hz; TN 200, 400, 600, 1000 Hz) and 400 (50% of all) presentations of the control (no tone) stimulus. The order of presentations was randomized, with the inter-stimulus interval between 500-2000 ms.

Procedure

The experiment lasted about 1 hour. Subjects sat in a quiet room in front of a PC computer, fixated on a white cross in the middle of the computer screen, and performed a tone detection task (50% of the trials) by pressing a mouse button, with the right index finger, as soon as they heard a tone popping out of the noise. Reaction times (RT) and accuracy scores were recorded. The participants were not informed about the existence of different types (HP vs. TN) of tonal stimuli. The presentation was divided into 4 blocks of 200 stimuli each. Between successive blocks, participants were allowed a short rest.

Before beginning the experiment proper, each listener was presented with a

short (approximately 5 minutes) practice session in which response feedback was provided on the computer screen. The purpose of the practice session was to familiarize the participants with the requirements of the task and to point their attention to the tonal objects, which because of their faintness sometimes are not detected without directed attention. If miss rates were high (more than 20%) the practice session was repeated but in no case were more than 2 sessions administered. Since a major experimental hypothesis was regarding dyslexic listeners' ability to detect HP stimuli, we chose to not present HP stimuli in the practice session (we reasoned that negative feedback in HP trials might discourage the listeners who are unable to perceive the tonal object). Instead of HP stimuli, we used noise band stimuli (BN). These signals are created using the same noise used to generate the HP stimuli but with one (narrow) band amplified. The amplified bands have the same bandwidth as the phase shifted region in the corresponding HP stimulus, but no interaural phase difference (i.e. they are audible monaurally). The loudness of the pitch in the BN stimuli was separately adjusted by 2 listeners to match the perceived loudness of the TN stimuli.

2.3.2 Results

Control subjects

The data for the control subjects is presented in Figure 2.19 (filled symbols). A repeated measures ANOVA with type (HP vs TN) and frequency as factors revealed a main effect of both (F(1,8)=11.56, p=0.009; F(3,24)=11.94, p<0.001) and no
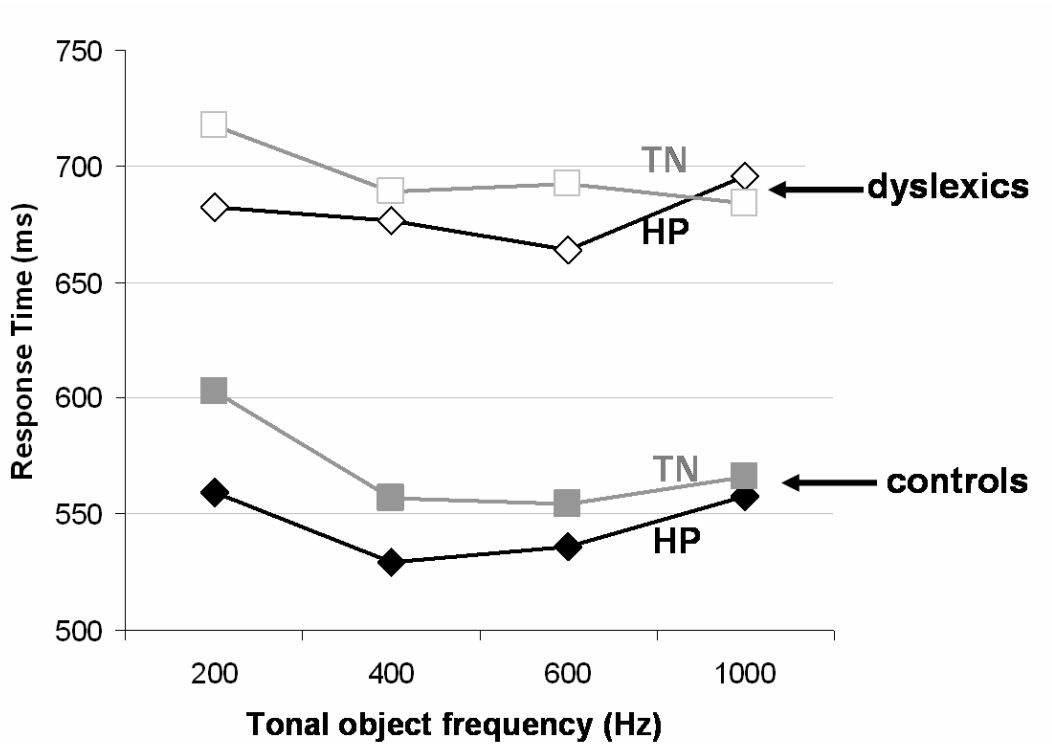
Figure 2.19: Response Times. HP: black lines, TN: grey lines. Controls: filled symbols, Dyslexics: empty symbols. Controls and dyslexics showed the same pattern of dependency on stimulus type (HP responses preceding TN responses), consistent with Chait et al. (2006). Dyslexic subjects performed similarly to controls apart from globally slower reaction times.

interactions. As in Chait et al (2006), subjects were consistently faster for HP than TN stimuli, with an average reaction time difference of approximately 30ms .The difference at 1000Hz is smaller than the one reported in (Chait et al, 2006) possibly due to group variance. In terms of miss rates, there was no difference between HP and TN conditions. On average, listeners missed about 6.9 targets (1.7%). The average false positive count was 19.7 targets (4.9%) with a standard deviation of 27.26.

Dyslexic subjects

The data for the dyslexic subjects is presented in Figure 2.19 (empty symbols). A repeated measures ANOVA with type (HP vs. TN) and frequency as factors revealed main effects of type ($F(1,6)=8.954$, $p=0.024$), freq ($F(3,18)=3.23$, $p=0.047$) and an interaction type×freq ($F(3,18)=6.543$, $p=0.003$). Subjects were consistently faster for HP than TN stimuli, with an average reaction time difference of approximately 30ms. As in the controls' data, the difference at 1000Hz is smaller than the one reported in (Chait et al, 2006) but this is a small effect.

Unlike the results of Dougherty et al (1998) it is clear from the present data that HP is audible to dyslexic listeners: the average number of HP misses (collapsed over frequency) is 12.71 (stdev = 7.76) out of 200 HP tokens (6.3 %). There was no difference in miss rates between HP and TN conditions. On average, listeners missed about 18.28 targets (4.5%). The average false positive count was 7 targets (1.7%) with a standard deviation of 10.64.
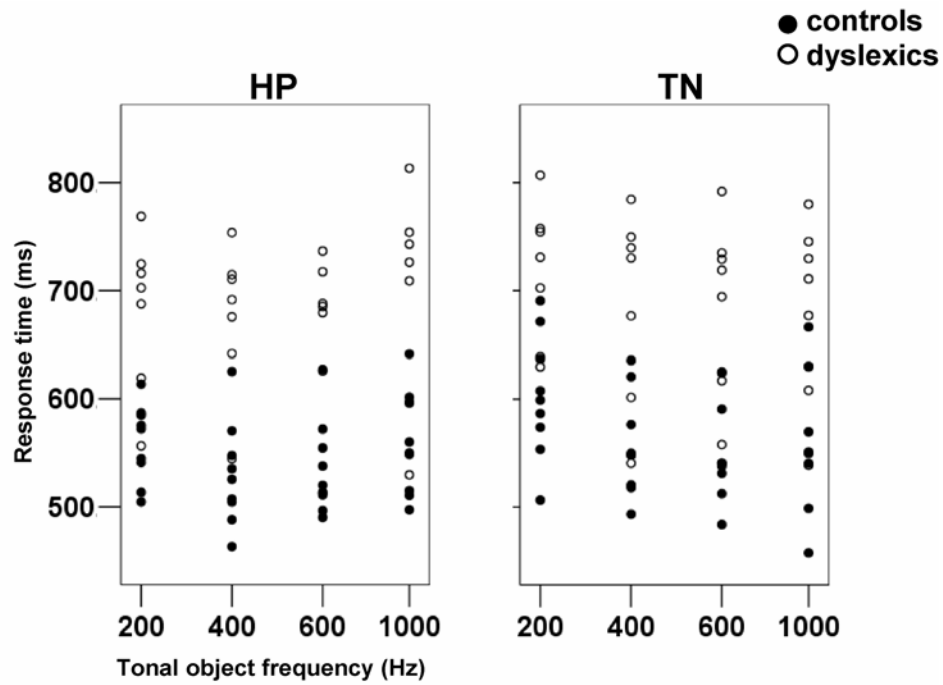
Figure 2.20: Scatter plots showing response time data for HP (left panel) and TN (right panel) for all our subjects. Controls: filled circles; Dyslexics: open circles. The dyslexic and control groups show little overlap except that two of our dyslexic subjects consistently perform in the same range as the controls. This is in line with a common observation that not all dyslexics show poor performance in auditory tasks (e.g Ramus, 2003).

Comparison between control and dyslexic data

Figure 2.20 shows the response times for all subjects in the control (black circles) and dyslexic (white circles) groups. The two subject groups exhibit almost no overlap. A repeated measures ANOVA for the response time data, with type (HP vs TN) and frequency as within subject factors, and group (dyslexic vs. control) as a between-subjects factor showed significant main effects of group ($F(1,14)=17.43$, $p<0.001$), type ($F(1,14)=18.463$, $p<0.001$) and freq ($F(3,42)=14.148$, $p<0.001$) and a type×freq interaction ($F(3,42)=9.751$, $p<0.001$). Both controls and dyslexics show a pattern of results (HP responses preceding TN responses) that is consistent with those obtained in Chait et al. (2006). Importantly, there was no interaction between any of the factors and group, indicating no difference between the performance of the control and dyslexic populations on these stimuli, except that dyslexic listeners' overall response times were much slower (a difference of approximately 120ms). Importantly, there is no difference in the false positives count between the two groups ($F(1,14)=1.334$, $p=0.267$). In the analysis of the miss data, a repeated measures ANOVA with type (HP vs TN) as a within subject factor and group as a between subject factor showed only a main effect of type ($F(1,14)=10.239$, $p=0.006$), indicating that in both groups, HP targets were generally missed more often than TN targets, and a main effect of group ($F(1,14)=9.15$, $p=0.009$): As compared to their normal reading controls dyslexic listeners show a significantly greater miss count. There were no interactions between group and type.
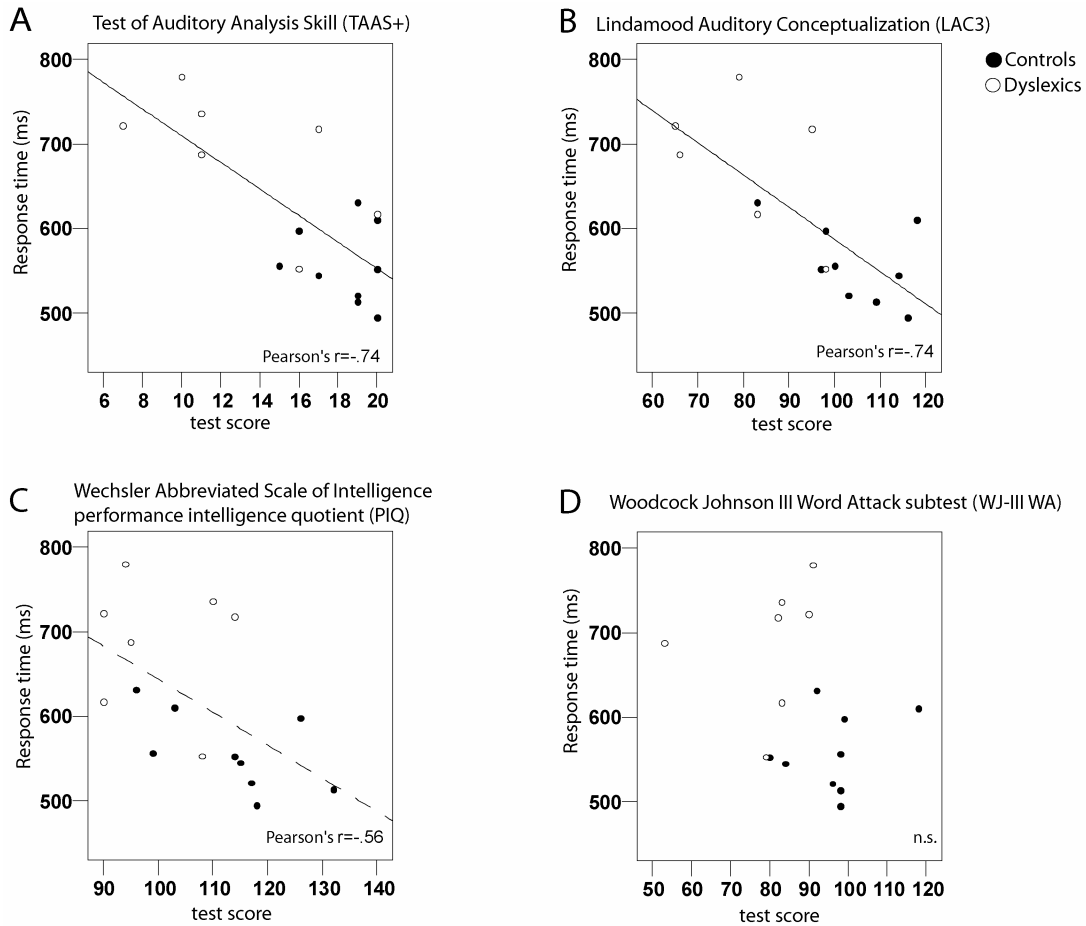
Figure 2.21: Scatter plots of test scores vs. response time. Controls: filled circles; Dyslexics: open circles. Black lines are linear fits to the data (R^2 values are 0.541, 0.552 and 0.316 in A, B and C respectively). Both auditory tests in our battery had a significant correlation with response times (A, B). There was also a correlation with the non-verbal IQ test (C), which further analysis showed to be a result of the distance between the two group means on both axes. All other common diagnostic tests for dyslexia, such as the WJ-III WA non-word reading test (D) are not correlated with the response time data.

We also compared the response variability of the two subject groups. A repeated measures ANOVA for the standard deviation data, with type (HP vs TN) and frequency as within subject factors, and group (dyslexic vs. control) as a between-subjects factor showed only a main effect of type $(F(1,14)=6.048, p=0.028)$ and no interactions. This effect was due to responses to HP stimuli being generally more variable than those to TN stimuli. Importantly, there was no main effect of group, suggesting that the response variability of dyslexic and control subjects did not differ.

We calculated Pearson correlations between mean response times (computed over all HP and TN conditions) and the scores in the test battery (see 'methods'). This analysis revealed significant strong correlations between response times and the two auditory tests in our battery: TAAS+ $(r=-0.736\ p=0.001)$ and LAC3 $(r=-0.743\ p=0.002)$ as well as, a substantially weaker, correlation with non-verbal IQ (PIQ; $r=-0.563\ p=0.023$). These results are presented in Figure 2.21 (5A-C). It is clear from the plots that the two fastest dyslexic subjects, who have response times similar to controls (see Figure 2.20), group with the controls in terms of LAC3 and TAAS+ test scores. On the other hand, the correlation with non-verbal IQ (figure 2.21C) does not exhibit that predictive power. Controlling for IQ (FSIQ parameter) and education reduce, but do not eliminate the correlation with LAC3 and TAAS+.

In order to examine to what extent the correlations are due to spread between the two populations, we subtracted the means from both the response times and test scores and computed the Pearson correlation again. The correlation with PIQ was found to be non significant under these conditions ($p=0.259$ n.s.), the correlation with

LAC3 was approaching significance ($p=0.07$), and the correlation with TAAS+ remained significant ($p=0.026$). This result implies that response times and the TAAS+ and LAC3 test scores are indeed linked, whereas the apparent correlation with PIQ test scores is a result of the separation of the means of the two subject groups.

2.3.3 Discussion

Our findings clearly demonstrate that dyslexic listeners can perceive dichotic pitch. None of our subjects had difficulty in identifying the HP targets (as exhibited by low miss and false positive counts). Moreover, the fact that the latency difference between TN and HP shown by the dyslexic listeners is not different than the latency difference exhibited by the control subjects indicates that in both populations the low level binaural integration mechanisms are operating in a similar manner, and that there is no specific binaural impairment associated with dyslexia.

Dyslexics were overall slower than controls by about 120 ms. Although we did not run a pure tone detection test in quiet (without background noise), the literature indicates that in simple reactions to pure tones, dyslexics respond as quickly as controls (e.g. Nicholson & Fawcett, 1994). Furthermore, the difference in miss-rates (but not false positive rates or response variability) between the subject groups suggest that the response time differences are not simply due to delayed motor execution or clumsiness, but are related to the processes via which the listeners detect the tonal objects in noise.

Because controls and dyslexics show a very similar pattern of dependency on stimulus type (apart from the latency difference), it is likely that the impairment resulting in increased response times is introduced at a processing stage in which TN and HP stimuli are already mapped to similar representations (i.e. above the level of the superior olive and IC, where binaural interaction occurs). Therefore, likely candidates would be thalamus or cortex.

That dyslexic listeners took more time than their controls to detect tonal objects in background noise is interesting because it appears to contradict a seemingly consistent experimental finding – usually there is no difference between the thresholds of dyslexics and controls in tasks involving the detection of long pure tones (similar to our TN stimuli) presented in the temporal center of a noisy background (Wright et al, 1997; McAnally & Stein, 1998; Hill et al., 1999; Amitay et al, 2002). We did not measure thresholds in this study, but the absence of difference in false positive rates and only small difference in miss rates between the subjects groups suggest that the tonal targets (both HP and TN) were similarly detectable. Thus, it would seem that the group difference is related to the *time* needed to process stimulus changes, rather than the ability to detect them - it was the *speed* with which the tonal targets were detected that underlies the impairment in dyslexic listeners. Threshold similarity tends to be interpreted as indicating normal auditory processing. Our results suggest that important aspects may escape threshold measures (especially when the detection threshold is set to be quite high), but may be probed by reaction-time measures.

What kinds of mechanisms could account for dyslexics' delayed responses? It may be that the deficit is specific to the processing of signals in noise. This interpretation is consistent with reports that reading impaired children show only subtle speech perception deficits in quiet but very serious difficulties under conditions of background noise (Ziegler et al., 2005). Similar findings, concerning deficits in noise exclusion in developmental dyslexia, have also been reported in the visual domain (Sperling et al., 2005). Another possibility is that dyslexic listeners are generally slow at tracking *changes* in ongoing stimulation – their change-detection integration windows are longer than normal -- and that this deficit is more evident amidst background noise. Indeed there is some evidence for a prolonged integration time in dyslexics (Helenius et al, 1999; Hari &Renvall, 2001; Hairston et al., 2005; Petkov et al, 2005). For simple stimuli such as the ones used here, a prolonged integration window would cause a delay in detection. In the case of more complicated signals, consisting of many consequent rapid changes such as speech, delayed change detection may cause increased inter-stimulus masking and result in a failure to properly perceive the sound sequence. This interpretation might also explain the apparent discrepancy between our results and those of Dougherty et al (1998), who employed signals similar to HP but reported significant impairment in individuals with dyslexia as compared to average readers. In their study, subjects listened to dichotic pitch melodies consisting of 4 sequential 200ms harmonic tone complexes. Inability to perform the task may have arisen from the longer time required by dyslexics to extract the tones from the noise and thus to follow the changes in the melody.

What might be the connection between the current findings and phonological processing? There is considerable controversy vis-à-vis the relationship of the variety of auditory deficits associated with dyslexics and their defining deficit, which is reading difficulty and phonological processing. One approach views dyslexia as a specific phonological disorder that sometimes co-arises with a more general sensory-motor syndrome (Bretherton & Homes, 2003; Ramus, 2004). According to this view, reading impairments are unrelated to low level auditory deficits (Rosen, 2003). Another major theory suggests the impairments associated with dyslexia are not speech specific but that phonological deficits arise from lower-level impairments in neural systems responsible for processing of auditory stimuli that have short duration and appear in rapid succession (Tallal, 1980; Tallal et al., 1993). Some researchers hypothesize that these auditory deficits as well as visual deficits related to rapid processing (Livingstone et al., 1991; Eden et al. 1996) arise from a general defect in the magnocellular processing pathways (Galaburda et al., 1985; Galaburda et al., 1994; Stein & Walsh, 1997; Stein, 2001) that sub-serve the processing of fast sensory (auditory, visual, tactile) information. A priori, our stimuli do not tap rapid auditory processing, but an account in terms of increased integration time in dyslexics and a consequential slower detection of changes (*any* change, not necessarily only rapid changes) might be able to explain much of the rapid processing deficits dyslexics seem to exhibit (e.g. Hari & Renvall, 2001). Hari & Renvall (see also Petkov et al, 2005) formulate their model in terms of 'sluggish attentional shifting' in dyslexia. This model suggests that sluggish attentional capture and prolonged attentional dwell result in an expanded 'cognitive integration window'. Actually, it is not clear that

attentional mechanisms need to be invoked in explaining increased integration time. Increased integration time and delayed detection may indeed arise from quite early 'pre-attentive' processing mechanisms (Banai et al., 2005; Chait et al, 2005).

It is noteworthy that, out of our extensive test battery, only tasks that rely on auditory analysis showed a significantly strong correlation with response times (Figure 2.21). The TAAS+ test is an auditory segmentation test of phonemic awareness. It asks subjects to remove a sound (beginning, ending or blend) from a spoken word and say what is left. For example, say 'bake' without the /b/ sounds (ache); say 'blend' without the /l/ sounds (bend). The LAC3, is a test that uses colored blocks to represent sounds. The subject is asked to manipulate non-words that are presented as sequences of these blocks (e.g., change blocks representing 'zap' to 'zup' by changing the middle block to another color). All other tasks involve visual stimuli and (except for performance IQ) require reading, naming, or verbal expression. For example, the WJ-III WA non-word reading test, on which the subject groups differed significantly at (Table 1) was not correlated with response time (Figure 2.21).

The strong correlations between the speed with which listeners detected the tonal objects and the auditory phonological processing tests may suggest that indeed both kinds of tasks rely on similar processing mechanisms. These findings are also consistent with previous reports of dissociation between speech processing impairments and reading skills (Agnew et al., 2004).Our sample sizes are too small, however, to draw strong conclusions.

In a previous MEG study with average-reading (non-dyslexic) subjects using the same stimuli and behavioral task as here (Chait et al, 2006), we identified a

121

cortical response, at approximately 160-200ms post tonal onset, evoked by the appearance of a tone (HP or TN) in the noise. We hypothesized that this increase in current is generated by the neural mechanisms that are responsible for the extraction of tonal objects from background noise. Several properties of this response are relevant to the current report: These 'tone onset' responses, localized to auditory cortex, were found to activate left hemispheric mechanisms significantly more strongly than right hemisphere ones. The left lateralization of the response was quite impressive and very robust (see figure 8 in Chait et al., 2006; figure 2.8 here). Interestingly, left temporal cortex is one of the brain regions that are most consistently reported to exhibit impaired activation in dyslexia (Eden & Zeffiro, 1998; Temple et al., 2000; Temple, 2002; Simos et al., 2002; Shaywitz et al 2004; Ramus 2004) and is also correlated with phonemic awareness skills (Turkeltaub et al., 2003). Thus, under-recruited neural substrate in the left temporal lobe may be a cortical reflection of difficulties exhibited by our dyslexic subjects.

Although the latency of these early MEG responses and behavioral reaction time followed similar general trends, they were not correlated (Chait et al, 2006; Chait & Simon, in press), which suggests that the tone onset responses reflect a pre-attentive stage in processing. This conclusion is consistent with other studies where similar responses were recorded even while subjects were not paying attention to the auditory stimuli (Krumbholz et al., 2003; Alain et al., 2002).   The pattern of early cortical response latencies vs. behavioral reaction times suggests that subjects' response times were also influenced by a later mechanism related to deciding, based on information from the earlier sensory mechanisms, whether the target has indeed

occurred (Chait et al, 2006; Chait & Simon, in press). These mechanisms are hypothesized to be the stage where the appearance of the target is brought to the subject's attention (Carpenter, 2004). Dyslexic subjects' delayed responses may thus have arisen from impaired sensory (auditory cortical or sub-cortical) processing, or from impaired higher level mechanisms. An MEG study with dyslexic listeners under similar listening conditions would clarify whether the impairments observed in the current study are indeed arising from early cortical mechanisms that underlie the extraction of the tonal objects, independent of attention, or are introduced by higher level mechanisms related to attention shifting and decision making (Hari &Renvall, 2001). Abnormal early cortical processing, as opposed to higher level impairment, is expected to be manifested in increased latencies of the early 'tone onset' responses, reflecting a delay in the neural mechanisms responsible for the segregation of the tonal targets from the noise. Normal 'tone onset' responses, but delayed behavioral detection, would implicate higher level (non-sensory) mechanisms.

Most previous brain imaging studies of dyslexia have focused on localization of impaired brain mechanisms (Eden et al., 1996; Eden & Zeffiro, 1998; Temple et al., 2002) or on investigating the 'fast temporal processing' hypothesis with response amplitudes as the main measure of detection thresholds (Nagarajan et al., 1999; Schulte-Körne et al., 1999; Kujala et al., 2000; Kujala et al., 2003). In light of the evidence provided here, that *latencies* of responses to simple auditory transients might reveal impairments in dyslexic listeners, future studies with stimuli such as the ones used here may shed light on the exact properties of the deficits observed in the present work.   Such an investigation may improve our understanding of dyslexia and,

possibly, of the computations involved in phonological processing. At the same time, and even if the sensory impairments are not causally related to the phonological deficits at the core of reading impairment, the investigation of auditory processing in dyslexia is extremely valuable for hearing research: The pattern by which the system 'breaks down' may serve as a fascinating means to further our understanding of the computations and neural mechanisms involved in the construction of perceptual representations.

2.4 Further Issues

In the series of experiments described above, we investigate the temporal dynamics of the process of extraction of a faint tonal target from background noise. We identify a cortical deflection, at approximately 150ms post target onset, which corresponds to the pre-attentive extraction of the target from the noise, but show that additional computations are involved in bringing it into consciousness. The weight of such conclusions, derived from a small number of experiments using a limited sampling of possible configurations and parameters, is obviously limited. The conclusions must be confronted with a wider range of situations. In this closing section of the chapter I will point out and discuss several outstanding issues.

2.4.1 Figure/ground segregation

Do the brain activations observed here reflect a general figure/ground extraction mechanism (e.g. such as the one suggested by Scheich et al, 1998) or are they somehow specific to our choice of paradigm and stimuli, for example the use of wide-band noise as a 'background'?

To address this question, we conducted another experiment, using pure tones as background (Figure 2.22). The task and procedure remained identical to the one used in the tone-in-noise experiments. Listeners (N=7) were presented with 1500ms long stimuli consisting of a 1500ms long loud pure tone (background; 990Hz) with

another, faint, tone (target; 200, 400, 600 or 800 Hz) appearing at 1000ms post onset. The loudness of the 'target' tone was adjusted by two listeners to be similar to the loudness of the TN stimuli in Chait et al (2006).

Figure 2.23 shows the group RMS of the auditory evoked responses for each of the conditions in the left and right hemispheres. The initial onset responses are discussed in chapter 1. The onset of the target tone evokes a vigorous response at about 150ms post target tone onset (1150ms post stimulus onset), with a characteristic M100 dipolar distribution. This is about 50ms earlier than the response to TN stimuli in Chait et al (2006)[11]. A thorough comparison of the responses found with the two 'backgrounds' would require a detailed source analysis (possibly using brain imaging techniques with better spatial resolution than MEG),and the present data are not compatible with such an analysis. However, even without source analysis, certain characteristics of the responses indicate that they arise from differing neural computations: Figures 2.24 and 2.25 present hemispheric activation comparisons for the stimuli of the two experiments. One of the strong results of Chait et al (2006) is that the tone onset response (for both HP and TN signals) activated the left hemisphere significantly more strongly than the right hemisphere (see figure 2.8 here). This observation is consistent with other reports of stronger left hemisphere activation for stimuli consisting of tones in noise (e.g. Hertrich et al., 2005; Hautus & Johnson, 2004; Johnson et al., 2003; Soeta et al, 2005).

In order to verify that the left lateralization is indeed consistent among the 20 listeners who participated in Experiment 1 of Chait et al (2006), we subjected the data to a bootstrap operation in which, at each iteration, 7 subjects were chosen in random

---

[11] For a possible explanation of this discrepancy, see Chapter 2.

and the corresponding hemispheric difference data was analyzed. Results showed that the left hemisphere advantage is indeed robust across subjects. Figure 2.24 plots the hemispheric activation data for a random group of 7 subjects from experiment 1 in Chait et al (2006). Figure 2.25 plots the hemispheric activation results for the 7 subjects in the follow up (tone background) experiment. The activation pattern in Figure 2.25 reveals that the tone-onset response in the 'tone background' condition exhibits markedly different properties – there is no strong left lateralization, and if anything an early right lateralization is visible. Procedures were identical in both experiments, and the bootstrap suggests that the difference in response pattern is not easily explained by the fact that the subjects in the latter experiment were few and different from those in the former experiment.

These results suggest that identical tonal targets may be processed differently, depending on the background from which they are extracted. Kubovy and van Valkenburg (2001), propose that auditory objects are defined in terms of the background from which they emerge. Another explanation may be that target detection requires different strategies in each case, although this account seems less likely because left lateralization is observed even when subjects are not actively paying attention to the tone-in-noise signals (Hertrich et al., 2005; Hautus & Johnson, 2004; Johnson et al., 2003; Soeta et al, 2005). Studies investigating listeners' ability to separate concurrent harmonic sounds report that successful extraction of a target sound is achieved by a suppression of the background rather than enhancement of the target (de Cheveigne', 1993; see also Lu & Dosher, 1998 for similar findings in the visual domain). The differing activation observed here may thus be related to

127

A　PRE stimuli

B　MAIN stimuli:

target stimuli

control stimuli

1000ms post onset

C　POST stimuli
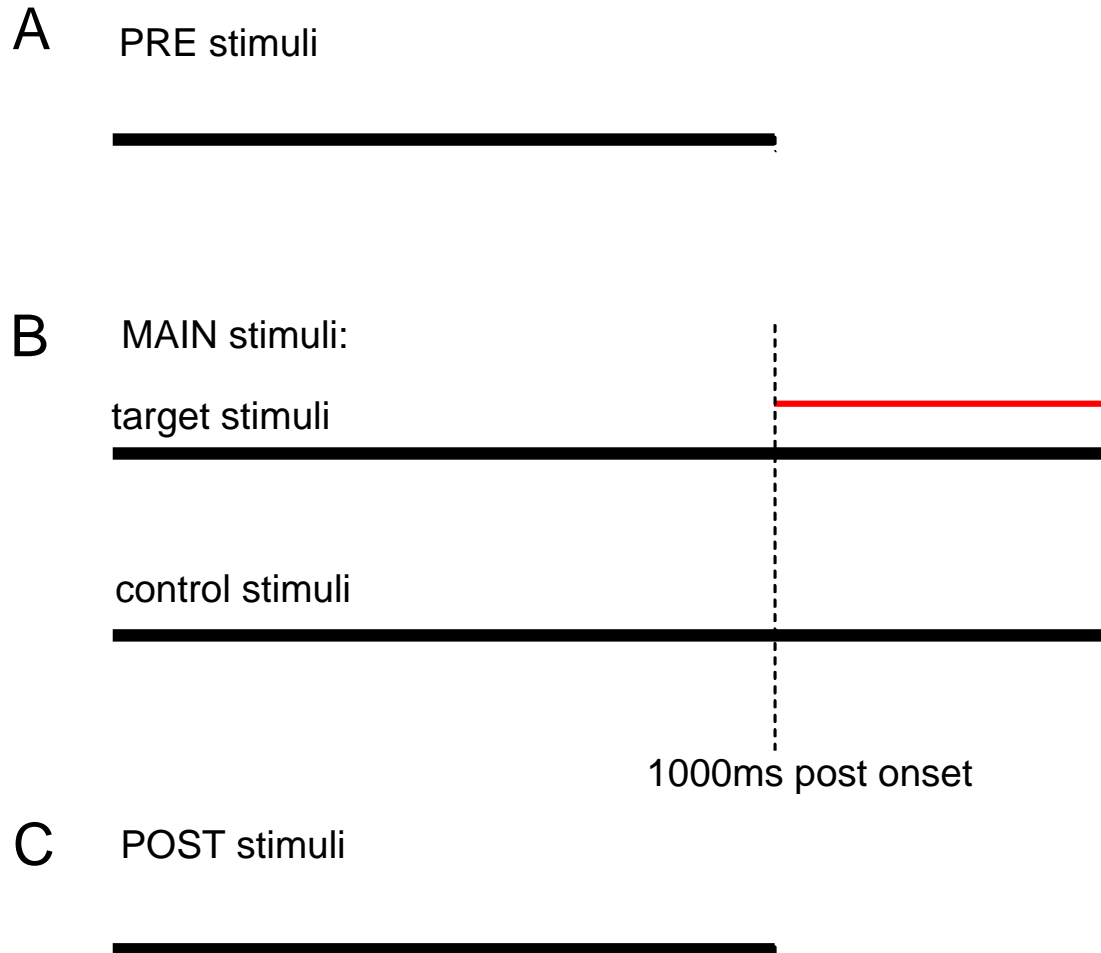
Figure 2.22. Schema of the stimuli used in follow-up experiment. The experiment consisted of two stages: (A) a pre-experiment where subjects listened to approximately 200 repetitions of a 1000ms long 990Hz pure tones and asked to count the signals, (B) Main experiment: Stimuli consisted of 1500ms 990Hz pure tone (black line) with a 500ms faint tonal object (red line) appearing at 1000ms post onset. The control stimulus was a 1500ms 990Hz pure tone. (C) At the end of the experiment subjects again listened to approximately 200 repetitions of 1000ms long 990Hz tones and asked to count the signals.
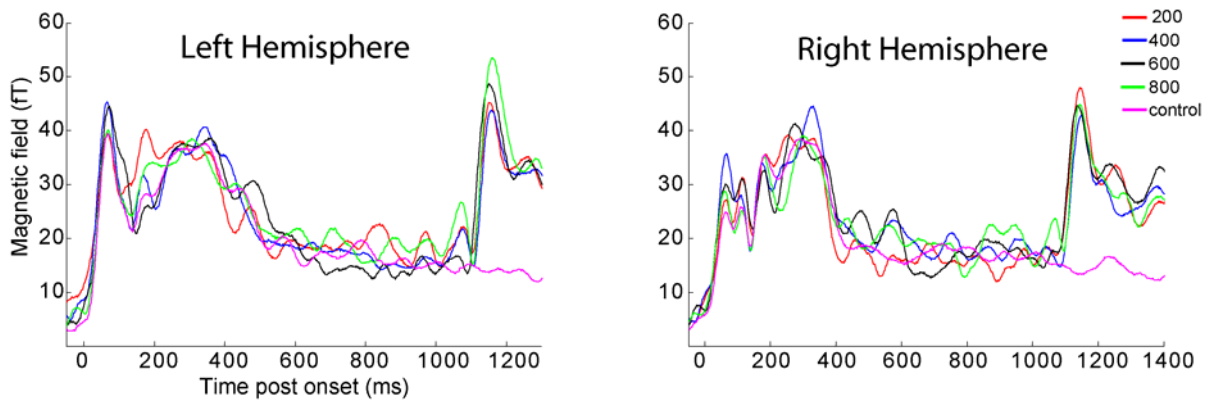
Figure 2.23. Summary of the electrophysiological data from the follow up experiment. The group-RMS in the LH and RH for all tested conditions. The response is characterized by an initial onset response (discussed in chapter 1), and a tone onset response at ~150 ms target onset.
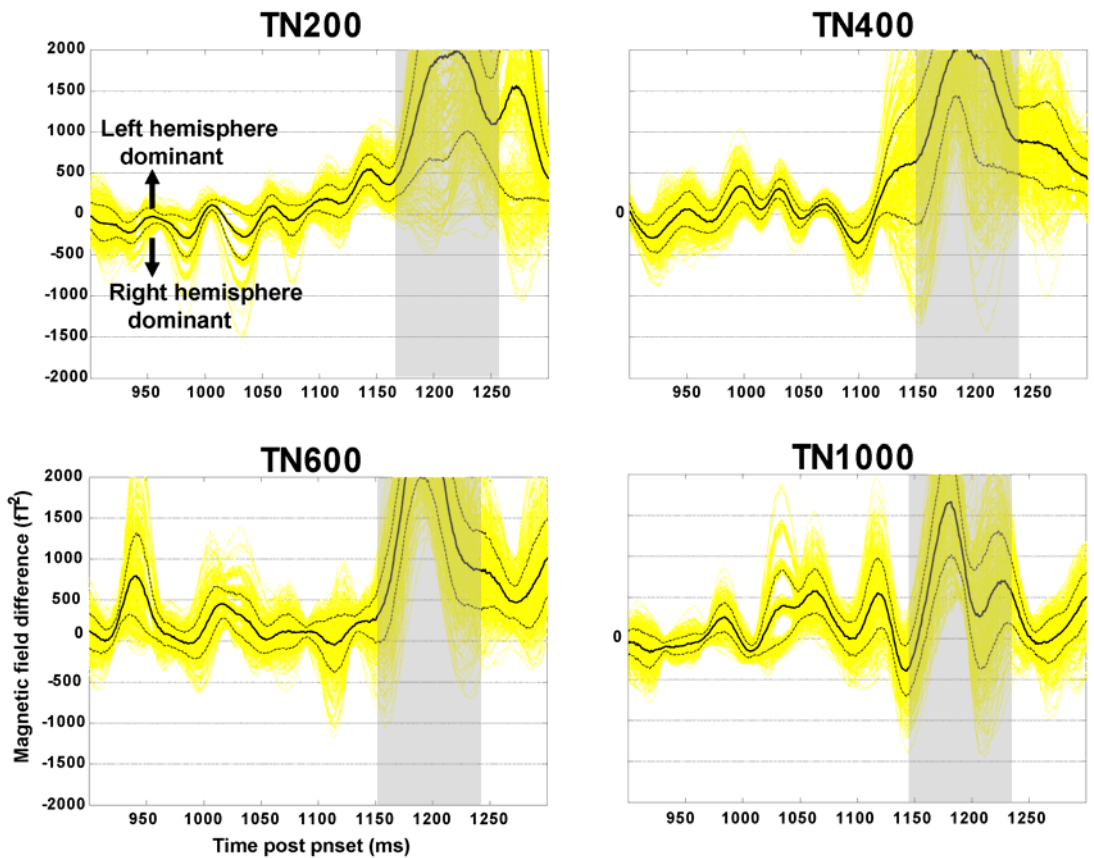
Figure 2.24. Comparison of hemispheric activation for TN stimuli in Experiment 1 of Chait et al (2006). Data for a random group of 7 participants. For each subject, right hemisphere activation was subtracted from left hemisphere activation and the result was bootstrapped across subjects. Yellow lines are the resulting bootstrap iterations (N=500). The solid black line is the bootstrap mean, and the dotted lines are one standard deviation. Tone onset responses for all conditions exhibited stronger left hemispheric activation.
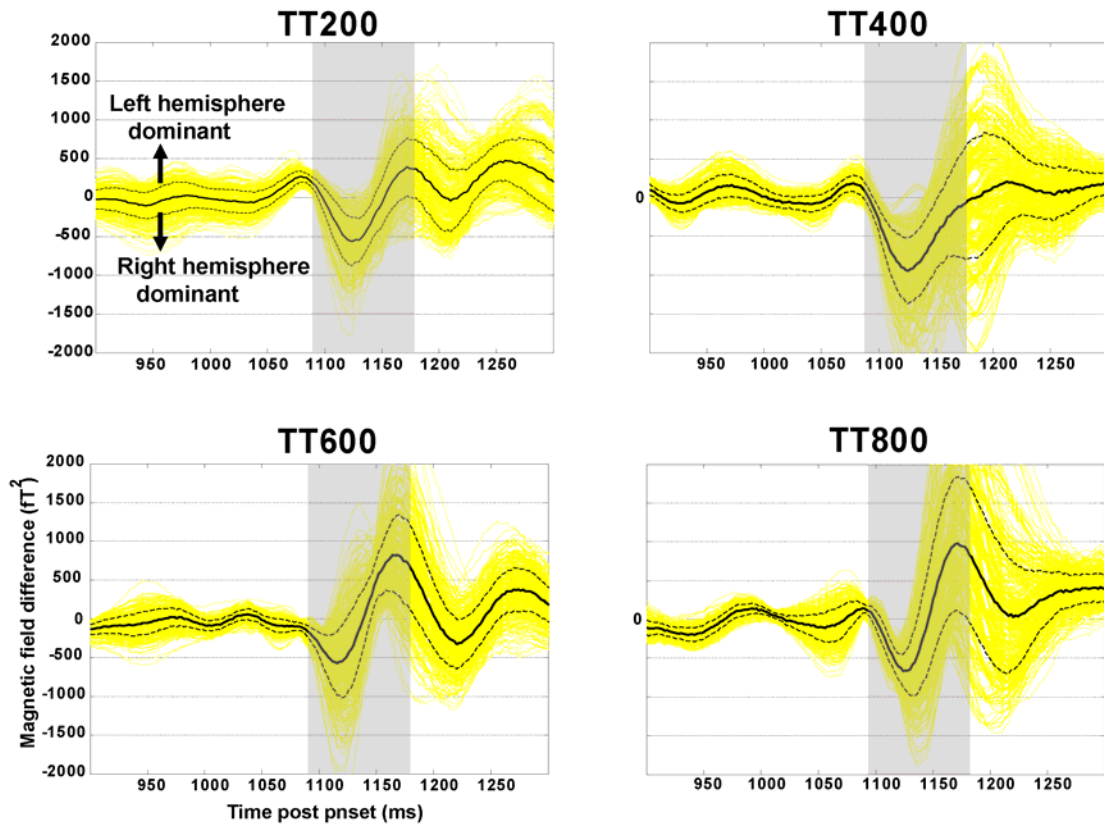
Figure 2.25. Comparison of hemispheric activation for the stimuli in the follow up experiment. For each subject, right hemisphere activation was subtracted from left hemisphere activation and the result was bootstrapped across subjects. Yellow lines are the resulting bootstrap iterations (N=500). The solid black line is the bootstrap mean, and the dotted lines are one standard deviation. The pattern of hemispheric activation differs from that of experiment 1 in Chait et al (2006). Strong left hemispheric activation is not observed, but most stimuli exhibit a pattern where the right hemisphere shows stronger initial activation.

different mechanisms employed in suppressing a background noise or a background tone.

This difference in outcome between seemingly similar experiments serves to alert us to the complexity of these phenomena, and the risk of hasty interpretation. It is worth noting in this context that in an additional experiment (N=10, described in Chapter 1), we presented both tone-in-noise and tone-in-tone stimuli in a randomized order. In this case, all hemispheric effects disappeared (see figure 1.3 here). Our current hypothesis to account for the lateralization discrepancies between experiments is that experimental outcome is strongly affected by the listening environment (stimulus set) and task (see chapter 1). We conclude that (a) it is extremely important to control for these parameters, even in cases where they appear irrelevant to the feature or task being investigated, (b) more experiments are needed to map out these dependencies.

2.4.2 Are changes in figure and ground processed separately?

Experiment 2 of Chait et al (2006), in which the onset of the tonal stimulus coincided with a change in interaural correlation of the noise background, suggested that tonal onset and noise change were processed by separate mechanisms, even though the change was concurrent in the physical signal. We argued that this result is potentially very interesting for auditory scene analysis research because it suggests that by ~130ms post change the noise and the tone already constitute separate 'auditory objects'. However such a strong conclusion may be unjustified if based solely on the

present results. In the current study *all* stimuli contained a change in the noise-background (see Figure 2.2). One possible explanation for the observed pattern of activation is that the repetition of the change in the noise in fact primed the relevant mechanisms resulting in a faster response. Note that this potential account of the data poses no problem for the conclusions discussed in sections 1 and 2 of this chapter. What was relevant for those arguments is our ability to separate the responses to the noise and responses to the tone (and not so much to explain *why* they are independent).

In order to investigate the possibility that the two changes are indeed reacted to separately by auditory cortex, an experiment has to be repeated where changes in the noise background and appearance of a tone are independently manipulated. This is also discussed at the end of chapter 3 here.

2.4.3 What computations underlie the tone-onset response?

Auditory objects may differ from the background in terms of power or frequency content. They may also differ in terms of the *statistics* (e.g. variance) of such quantities. In order to detect that an auditory object has appeared, a listener must monitor the statistical properties of the ongoing sound and react when it detects a change. A variation unexpected on the basis of the natural fluctuations of existing sound sources in the environment, would indicate that a new object has appeared in the environment. In this sense, object detection and change detection are synonymous. In the next chapter, I will demonstrate that the properties of the

responses observed here are indeed interpretable in the framework of change detection.

Our experiments with dyslexic subjects add an additional dimension to this inquiry. The finding that individuals with dyslexia (or at least a subset of them) are impaired at detecting tonal objects in background noise is very interesting. If our future experiments, in particular with MEG, show that the deficit stems from the mechanisms underlying the tone-onset response, this will open the possibility of studying more closely the processing that follows change detection. and what consequence impairment of this processing may have on how perceptual representations are constructed. As reviewed above, several auditory deficits associated with dyslexia may be re-interpreted in the frame work of impaired change detection. I view this as a very exciting avenue for the future.

Chapter 3: Change detection

Sensitivity to changes in sound is important to auditory scene analysis and detection of the appearance of new objects in the environment. In this chapter I describe two separate experiments which investigated the temporal dynamics of the process by which listeners detect changes in ongoing stimuli. In one experiment, that associated behavior and neurophysiological recordings, we studied changes in the interaural-correlation (IAC; the degree of similarly of the signal at the two ears) of wide-band noise. Stimuli consisted of interaurally correlated noise (identical noise signals played to the two ears) that changed into uncorrelated noise (different noise signals at the two ears) or vice versa. IAC is a basic cue for binaural processing, and mechanisms that are sensitive to changes in IAC have been implicated in processes subserving auditory scene analysis (e.g. binaural unmasking; Durlach et al., 1986).

The second experiment also associated behavior and brain recording. Stimuli were designed to mimic the abstract properties of those in the IAC experiment, while changing the acoustic properties completely. Signals consisted of a constant tone that changed into a sequence of random tone pips, or vice versa. As we shall see, the outcome of these two experiments was remarkably similar. The stimuli of the two experiments differ acoustically (broadband, binaural, and stationary vs. narrowband, monaural and dynamic), and the perceptual attributes evoked by the change are different. Nevertheless, at an abstract level they share the same characteristic of a transition between 'order' and 'disorder'.

Transitions such as these are relevant to auditory scene analysis: the appearance of a coherent object within noisy background can be understood as a sudden emergence of order within the ongoing random signal. Likewise, the appearance of an unexpected event can be understood as a disruption of a previous, predictable, order. Figures 3.1 and 3.2 illustrate the stimuli in each of the experiments and allow one to appreciate their conceptual similarity. We will interpret the similarity of their outcome as reflecting the operation of a general change mechanism, that handles in a common fashion changes along very different stimulus statistics dimensions.

The first experiment (IAC) is described in Section 1 where I reproduce Chait et al (2005) in its published form. The second experiment (tone sequences) is reported in Section 2. Implications of the findings and further issues are discussed in Section 3.
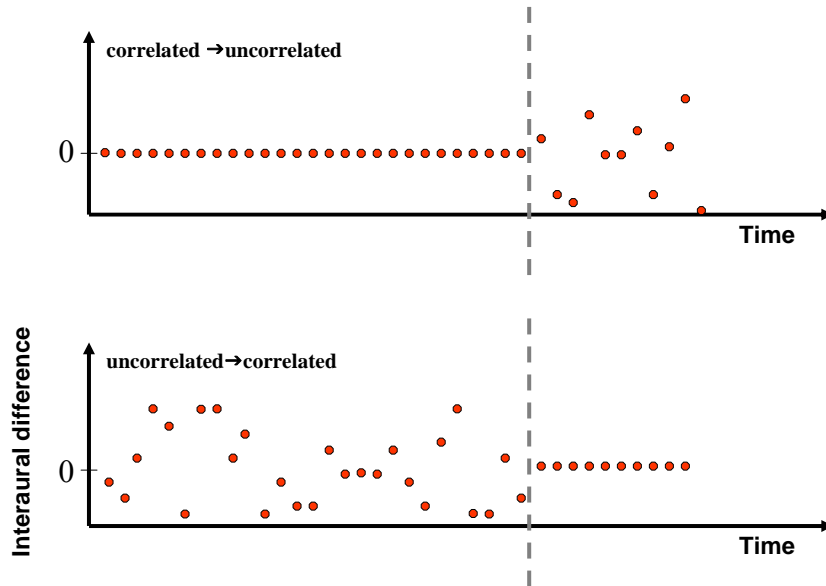
Figure 3.1: Schema of the stimuli used in the interaural correlation experiment. The dimension of change (Y axis) is interaural difference. We measured the cortical and behavioral sensitivity of listeners in detecting transitions from correlated to uncorrelated noise (top) and vice versa (bottom). Correlated noise is hypothesized (Durlach, 1963) to be represented as a constant "zero" value, whereas uncorrelated noise is represented as a randomly fluctuating value. The size of the red dots symbolizes the temporal resolution of binaural processing.
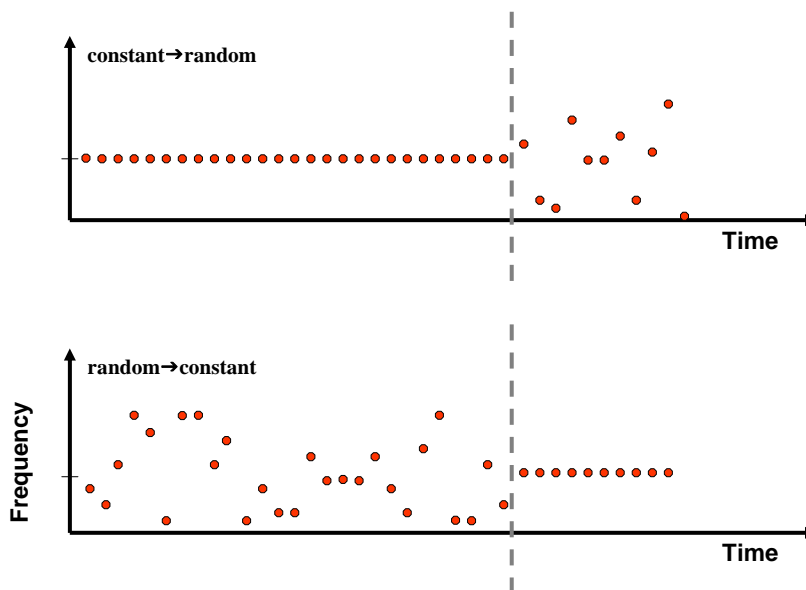
Figure 3.2: Schema of the stimuli used in tone change detection experiment. The dimension of change (Y axis) is frequency. We measured the cortical and behavioral sensitivity of listeners in detecting transitions from a constant tone to a random sequence of tones (top) and vice versa (bottom). The size of the red dots symbolizes the size of a tone-pip (we used 60 ms, 30ms and 15ms tone pips and separate blocks of the experiment).

3.1 Human auditory cortical processing of changes in interaural correlation[12]

Ecologically relevant tasks, such as detection and localization of auditory objects in noisy environments, involve comparison of acoustic signals across ears. Interaural coherence - the degree of similarity of the waveforms at the two ears, is a basic cue for binaural processing. In addition to being closely related to the mechanisms that underlie the localization of sound (e.g. Stern & Trahiotis, 1995), the detection of a change in the interaural coherence of an ongoing background is thought to be the primary cue in situations where binaural unmasking occurs: a target that is masked by binaurally correlated noise (identical noise at the two ears) can be made easier to detect by inverting the noise or the target in one ear (Hirsch, 1948; Licklider, 1948). Binaural unmasking is fundamental to listeners' ability to operate in noisy, multi-source environments and has been widely investigated both electrophysiologically (Jiang et al., 1997ab; Palmer et al., 2000) and behaviorally (see Colburn, 1995 for review). This phenomenon may be mediated by the auditory system's ability to detect decreases (in case of inverting the target) or increases (in case of inverting the noise) in interaural coherence resulting from the addition of the target (Durlach et al., 1986; Palmer et al., 1999). Therefore the investigation of the neural mechanisms that are sensitive to interaural similarity is particularly informative in the study of how listeners analyze the auditory scene and react to changes in the order of the

---

[12] This section is a reproduction of Chait M, Poeppel D, de Cheveigne A, Simon JZ (2005) Human auditory cortical sensitivity to changes in interaural correlation. J Neurosci 25:8513-27.

environment.

A physical measure of coherence is 'interaural correlation' (IAC), defined as the cross-correlation coefficient of the signals at the two ears. Several behavioral studies have measured listeners' ability to discriminate interaural correlations (Pollack & Trittipoe, 1959a,b; Gabriel & Colburn, 1981; Culling et al., 2001; Boehnke et al., 2002). Just noticeable differences (JNDs) are not uniform across the IAC range: they are small (typically 0.04) when measured as differences from an IAC value of 1 and an order of magnitude larger when measured as differences from IAC value of 0. Listeners are thus more sensitive to deviations from similarity than deviations from dissimilarity, at least as measured in terms of interaural correlation. It is unclear, however, at which level in the processing stream, from brainstem where information from the two ears is first merged, up to cortex where behavioral responses are initiated, this distinction is introduced.

Natural environments are characterized by dynamic changes in interaural correlation as objects appear and disappear. Here we combine, for the first time, psychophysical measures and non invasive brain-imaging via Magnetoencephalography (MEG) to study how the human auditory cortex processes these changes. Specifically, we measure early (~50-150 ms post change) cortical responses to changes in interaural coherence, and compare these to behavior. With its fine temporal resolution, MEG is particularly useful for studying the *time-course* of cortical activation, thus allowing for comparison with the time-course of behavioral responses and an investigation of the dynamics of the construction of perceptual experiences.

3.1.1 Materials and Methods

Subjects

Eighteen subjects (mean age 21.9 years, 11 female), took part in the MEG experiment. Fifteen subjects (mean age 21.9, 8 female) took part in the behavioral study. Ten listeners participated in both experiments. Three additional participants in the MEG study and one additional participant in the behavioral study were excluded from analysis due to an excess of non-neural artifacts in the MEG data or inability to perform the task. All subjects were right handed (Oldfield, 1971), reported normal hearing, and had no history of neurological disorder. The experimental procedures were approved by the University of Maryland institutional review board and written informed consent was obtained from each participant. Subjects were paid for their participation.

Stimuli

MEG: The signals were 1100 ms-long wide-band noise bursts, consisting of an initial 800 ms-long segment (reference correlation) that was either interaurally correlated (IAC = 1) or interaurally uncorrelated (IAC = 0), followed by a 300 ms segment with one of six fixed values of IAC: 1.0, 0.8, 0.6, 0.4, 0.2, 0.0. Human listeners' performance on detecting changes in IAC remains approximately constant for signal durations greater than 300 ms (Pollack & Trittipoe, 1959b). The purpose of the relatively long initial segment was to ensure that responses to change in IAC do

not overlap with those associated with stimulus onset. The bandwidth and spectral power were equal at each ear and constant across conditions. All signals sound the same when presented monaurally, and the change at 800 ms occurred without any detectable change in either monaural signal. Thus any differences in behavioral or brain responses can be interpreted as specifically resulting from binaural interaction.

Previous behavioral studies (Pollack & Trittipoe, 1959a,b; Gabriel & Colburn, 1981; Culling et al., 2001) suggest that equal IAC steps do not map to equal perceptual distance. In fact, the IAC scale that defines approximately equal perceptual steps has been suggested to be exponentially shaped, such as the scale (1.0, 0.93, 0.80, 0.6, 0.33, 0.0) which was used by Budd et al. (2003) in an fMRI study of static interaural correlation sensitivity. We chose to use a linear scale here in order to examine to what extent the perceptual non-linearity would be reflected in early cortical responses and whether we might observe dissociations between these neural representations and behavior. For that reason, for instance, it was interesting to measure brain responses to the $0 \rightarrow 0.2$ condition, for which the change is behaviorally unnoticeable. The choice of the "physical dimension" among various non-linearly related forms is arbitrary, and thus "non-linearity" of the function relating it to responses is not of interest per se. The form of the function is nevertheless worth investigating. Our choice of equally spaced IAC values determines the sampling of this function, but does not prejudge its shape.

The noise waveforms were constructed using the same paradigm as in Gabriel and Colburn (1981). Two independent 800 ms signals, denoted below as $n_1(t)$ and $n_2(t)$, were created by drawing Gaussian distributed numbers (sampling frequency 16

kHz). The signals presented to the left and right ears ($n_L(t)$ and $n_R(t)$, respectively) were constructed by mixing $n_1(t)$ and $n_2(t)$ according to the equations:

$$n_L(t) = n_1(t)$$

$$n_R(t) = \beta n_1(t) + \sqrt{(1-\beta^2)} n_2(t)$$

where β = 1.0, 0.8, 0.6, 0.4, 0.2 or 0.0. For exactly orthogonal $n_1(t)$ and $n_2(t)$ the interaural correlation coefficient of $n_L(t)$ and $n_R(t)$ is equal to the value of β (e.g. Gabriel & Colburn, 1981). In order to reduce response dependency on a particular sample of frozen noise, 10 different instances were generated for each of the 12 conditions. Due to the fact that random samples of noise are not exactly orthogonal, the value of the interaural correlation coefficient between $n_L(t)$ and $n_R(t)$ may differ slightly from its nominal value β. The standard deviation of the difference averaged over all conditions was 0.009.

In addition to the 12 experimental conditions, the stimulus set included a proportion (25%) of "target" (decoy) stimuli which consisted of 800 ms of either interaurally correlated (IAC = 1) or interaurally uncorrelated (IAC = 0) wide-band noise, followed by 300 ms of interaurally correlated (IAC = 1) or interaurally uncorrelated (IAC = 0) noise modulated at a rate of 10 Hz and a depth of 50%. Subjects were instructed to respond as fast as they could to each onset of the modulation. The target stimuli were not included in the analysis. Because of their high similarity to the experimental conditions they served to assure the subjects' alertness and to focus attention on the time of change (800 ms post onset) but did not require any conscious processing of interaural correlation.

The decoy task did not involve IAC processing to avoid influencing brain responses

to the main conditions. Decoy and main conditions were kept distinct to ensure that the MEG responses probe low level auditory processes and not higher level processes engaged by the task.

The stimuli were created off-line, gated on and off using 15 ms cosine-squared ramps (with no gating at the transition at 800 ms post onset), and saved in 16-bit stereo wave format at a sampling rate of 16kHz. The signals were delivered to the subjects' ears with a tubephone (E-A-RTONE 3A 50 ohm, Etymotic Research, Inc) attached to E-A-RLINK foam plugs inserted into the ear-canal and presented at a comfortable listening level.

In total each listener heard 120 repetitions of each of the 12 experimental conditions (0→0, 0→0.2, 0→0.4, 0→0.6, 0→0.8, 0→1, 1→0, 1→0.2, 1→0.4, 1→0.6, 1→0.8, 1→1) and 120 repetitions of each of the 4 target conditions (0→modulated 0, 0→modulated 1, 1→modulated 0, 1→modulated 1). The order of presentation was randomized, with the inter-stimulus interval (ISI) randomized between 600-1300 ms.

Perceptually, correlated noise (IAC = 1) sounds like a single focused source in the center of the head. The image broadens as interaural correlation decreases and at IAC = 0 the percept is that of a diffuse source, or two independent sources, one at each ear. Thus, 0→ stimuli evoke a percept of focusing of the sound image, whereas 1→ signals evoke a broadening of the source. The stimuli used in this study are illustrated by their binaural cross correlograms in Fig. 3.3. The correlograms were generated using the "binaural toolbox" (Akeroyd, 2001). In order to simulate peripheral processing, the acoustic signal of each ear (the 300ms long post change segment) was fed through a filter-bank (100 Hz to 2000 Hz with filter spacing of one

half an Equivalent Rectangular Bandwidth, Moore and Glasberg, 1983) and half wave rectified; left and right filter outputs were delayed, cross-multiplied, and normalized by the average power in the two filter outputs. Correlated noise is characterized by an orderly arrangement of "valleys" and "ridges" while uncorrelated noise evokes an irregular pattern with low amplitude. Decreasing values of IAC are characterized by a progressive fading of the valley/ridge structure and reduction of amplitude. Physiological evidence indicates that Medial Superior Olive (MSO) neurons are tuned to a particular input frequency (the characteristic frequency of the cell) and interaural time difference (ITD). Binaural models commonly approximate the MSO to an array of cross-correlators fed from both ears (Jeffress, 1948; Joris et al., 1998) and the plots in the figure illustrate the long-term (300ms) time average of the activity within such an array that would be evoked by our stimuli. Differential activation in the MSO may be the source of the differential activation we describe in auditory cortex, as discussed in detail below. The patterns illustrated in Fig. 3.3 reflect a generic crosscorrelation model, but a similar account could be applied to the recent model of McAlpine and colleagues (McAlpine & Grothe, 2003)

Behavioral study: The stimuli for the behavioral study were identical to the MEG stimuli except that the amplitude modulated, decoy, stimuli were not included. Instead, the proportion of stimuli without IAC change was increased to equal that with change. Each participant was given 300 presentations of each of the no-change conditions ($0 \rightarrow 0$, $1 \rightarrow 1$), and 60 presentations of each of the change conditions ($0 \rightarrow 0.2$, $0 \rightarrow 0.4$, $0 \rightarrow 0.6$, $0 \rightarrow 0.8$, $0 \rightarrow 1$, $1 \rightarrow 0$, $1 \rightarrow 0.2$, $1 \rightarrow 0.4$, $1 \rightarrow 0.6$, $1 \rightarrow 0.8$). Subjects were instructed to press a mouse button as fast as they could when they heard a
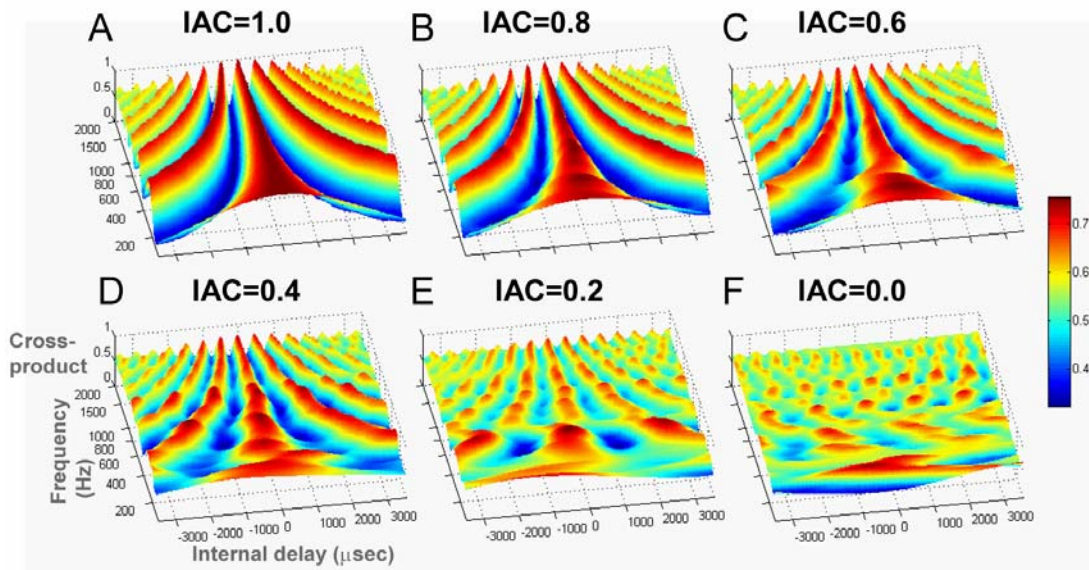
Figure 3.3: Binaural cross-correlograms for the stimuli used in this study (computed over the 300 ms-long post-change interval). Correlated noise (IAC = 1; A) is characterized by a systematic arrangement of peaks and troughs (main peak at zero ITD with side peaks spaced according to reciprocal of the center frequency in each band). Uncorrelated noise (IAC = 0; F) evokes an irregular pattern with low amplitude. Intermediate levels of IAC show a progressive reduction of amplitude and waning of the valley/ridge structure.

change in the noise. The order of presentations was randomized, with ISIs as in the MEG experiment.

Procedure

MEG: The subjects lay supine inside a magnetically shielded room. In a pre-experiment, ran just before the main experiment, subjects listened to 200 repetitions of a 1 kHz 50 ms sinusoidal tone (ISI randomized between 750-1550 ms). These responses were used to verify that signals from auditory cortex had a satisfactory signal to noise ratio (SNR), that the subject was positioned properly in the machine, and to determine which MEG channels best respond to activity within auditory cortex. In the experiment proper (about 1.5 hours), subjects listened to stimuli while performing the modulation detection task as described above. They were instructed to respond by pressing a button, held in the right hand, as soon as they heard a modulation appear in the noise. The instructions encouraged speed and accuracy. The experiment was divided into blocks of 160 stimuli. Between blocks subjects were allowed a short rest but were required to stay still.

Behavioral study: The experiment lasted about 1 hour. Subjects sat in a quiet room and were instructed to press a button held in their right hand as soon as they detect a change in the reference noise. No feedback was provided. Response-times and accuracy-scores were stored and analyzed. The experiment was divided into blocks of 200 stimuli. Between blocks subjects were allowed to rest but prohibited from getting up or removing the ear pieces. Before the experiment proper, subjects completed a short practice run with feedback. The stimulus delivery hardware, software, and headphones were identical to those used in the MEG recording.
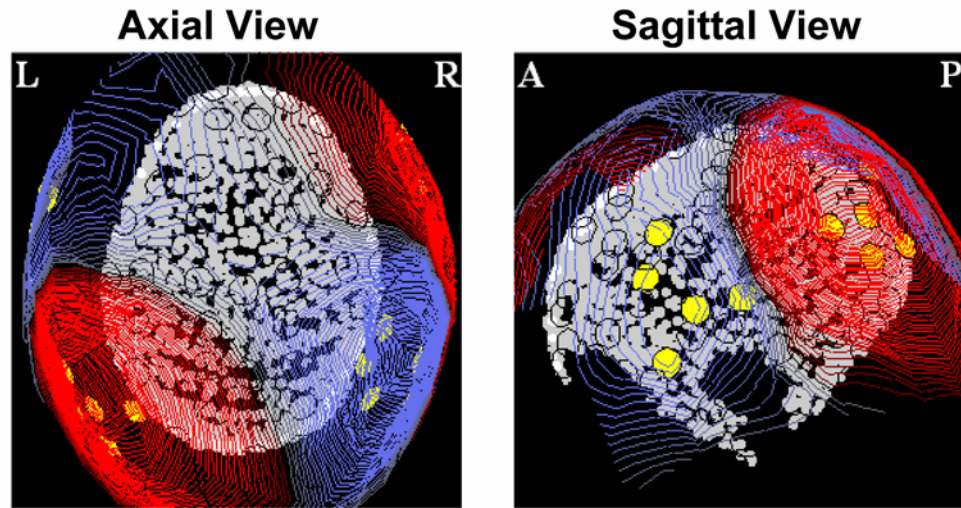
Figure 3.4: Dipolar pattern corresponding to the M100 responses in the pre-experiment. The figure shows an axial view (L = left, R = right) and a sagittal view (A = anterior, P = posterior) of the LH of the digitized head-shape of a representative subject. The magnetic dipole-like pattern in the iso-field maps distributed over the temporal region (red = source; blue = sink), corresponds to a downward flowing current. M50 responses are characterized by an opposite sink-source pattern (current flowing upward). Different channels were chosen for each individual subject depending on their M100 response in the pre-experiment; the locations of the 20 chosen channels for this subject are marked with yellow circles.

Neuromagnetic recording and data analysis

In this study we are particularly interested in the temporal characteristics of the brain responses evoked by our stimuli. These responses are contaminated by sensor noise, environmental fields, and brain activity unrelated to auditory processing. Several steps are taken to reduce this variability: (1) At each sensor, the response is partitioned into 'epochs' (including a short prestimulus interval) and averaged over repetitions. (2) Responses are high-pass filtered to remove slow baseline fluctuations in the magnetic field, and low-pass filtered to attenuate the (typically non-evoked) high-frequency components. (3) Measures are derived from a subset of sensors selected for each subjects (10 for each hemisphere) known to respond strongly, based on responses in the pre-experiment, to activity in auditory cortex. (4) The same measures are averaged over subjects, and the significance of effects is tested (independently for each hemisphere) by comparison to intersubject variability (repeated-measures analysis). Two measures of dynamics of cortical processing are reported: the amplitude time course (increases and decreases in activation) as reflected in the root mean square (RMS) of the selected channels, and the accompanying spatial distributions of the magnetic field (contour plots) at certain times post onset. For illustration purposes, we plot the group-RMS (RMS of individual RMSs, computed on the basis of the channels chosen for each subject) or the grand-average (average over all subjects for each of the 160 channels).

The magnetic signals were recorded using a 160-channel, whole-head axial gradiometer system (KIT, Kanazawa, Japan). Data for the pre-experiment were

acquired with a sampling rate of 1kHz, filtered online between 1 Hz (hardware filter) and 58.8 Hz (17 ms moving average filter), stored in 500 ms (100 ms pre-onset) stimulus-related epochs and baseline-corrected to the 100 ms pre-onset interval. Data for the main (interaural correlation) experiment were acquired continuously with a sampling rate of 1 kHz, filtered in hardware between 1 and 200 Hz with a notch at 60 Hz (to remove line noise), and stored for later analysis. Effects of environmental magnetic fields were reduced based on several sensors distant from the head using the CALM algorithm (Adachi et al., 2001), and responses were then smoothed by convolution with a 39ms Hanning window (cutoff 55 Hz). These are standard signal processing methods; further processing is described below.

In the pre-experiment, auditory evoked responses to the onset of the pure tones were examined and the M100 response identified. The M100 is a prominent and robust (across listeners and stimuli) deflection at about 100 ms post onset, and has been the most investigated auditory MEG response (see Roberts et al, 2000 for review). It was identified for each subject as a dipole-like pattern (i.e. a source/sink pair) in the magnetic field contour plots distributed over the temporal region of each hemisphere. In previous studies, under the same conditions, the resulting M100 current source localized to the upper banks of the superior temporal gyrus in both hemispheres (Hari, 1990; Pantev et al, 1995; Lütkenhöner & Steinsträter, 1998). For each subject, the 20 strongest channels at the peak of the M100 (5 in each sink and source, yielding 10 in each hemisphere) were considered to best reflect activity in the auditory cortex and thus chosen for the analysis of the experimental data (Fig. 3.4).

Stimulus-evoked magnetic fields, measured outside the head by MEG, are

149

generated by synchronous neuronal currents flowing in tens of thousands of cortical pyramidal cells on the supratemporal gyrus (Hämäläinen et al., 1993). This electromagnetic fluctuation is detected as a magnetic dipole with position, orientation and strength. Because of the location of the source inside a cortical fold, responses from auditory cortex typically manifest a characteristic dipolar distribution (source/sink pairs that are anti-symmetric across the two hemispheres). Fig. 3.4 shows a 3-D image of the dipolar pattern corresponding to the M100 responses in the pre-experiment. Later figures plot the same information in flattened 2-D contour maps.

In the main experiment, 1400 ms epochs (including 200 ms pre onset) were created for each of the twelve stimulus conditions. The same data were also organized into two additional compound conditions by grouping together all epochs with a reference correlation of 1 and 0, to improve the SNR of onset responses to correlated and uncorrelated sounds respectively. Epochs with amplitudes larger than 3 pT (~5%) were considered artifactual and discarded. The rest were averaged, low-pass filtered at 30 Hz (67 point wide Hanning window), and base-line corrected to the pre-onset interval. In each hemisphere, the root mean square (RMS) of the field strength across the 10 channels, selected in the pre-experiment, was calculated for each sample point. Twenty- eight RMS time series, one for each condition in each hemisphere, were thus created for each subject.

To evaluate congruity across subjects, the individual RMS time series were combined into twenty-eight group-RMS (RMS of individual RMSs) time series. Consistency of peaks in each group-RMS was automatically assessed with the Bootstrap method (500 iterations; balanced; Efron & Tibshirani, 1993). The

consistency, across subjects, of magnetic field distributions at those peaks was assessed automatically by dividing the 20 channels chosen for each subject into 4 sets (5 channels each): left temporo-frontal, left posterior-temporal, right temporo-frontal, right posterior-temporal (see Fig. 3.4). For each set, the activation was averaged over a 30ms window defined around the group-RMS peak, and the set was classified as either a "sink" (negative average amplitude) or a "source" (positive average amplitude). If the majority of subjects showed the same sink-source configuration, the pattern was considered consistent across subjects.

The α level for the statistical analyses was set *a priori* to 0.05. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied where applicable.

3.1.2 Results

Behavioral data

Accuracy scores and response times are summarized in Fig. 3.5. Our task differed from other studies in that our subjects had to detect a *transition* from an initial IAC of either 1 or 0, rather than a *difference* of IAC between temporally separate segments of noise presented in random order (Pollack & Trittipoe, 1959; Gabriel & Colburn, 1981; Culling et al, 2001). Nevertheless detection rates followed a similar trend (Fig. 3.5A). An ANOVA (over the change conditions) revealed main effects of reference correlation ($F(1) = 193.167\ p < 0.001$) and size of IAC step ($F(4) = 100.681\ \ p < 0.001$), as well as an interaction between these two factors ($F(4) = 48.73\ \ p < 0.001$). Subjects were good at detecting changes from an initial correlation

of 1 ("1→") but less good at detecting changes from an initial correlation of 0 ("0→").

In both cases (1→ and 0→) detection improved with the size of the IAC step between

the initial and final segments. Fig. 3.5B shows the corresponding response times.

Similarly to the detection rates, there were main effects of reference correlation ($F(1)$

$= 44.93$ $p < 0.001$) and size of IAC step ($F(4) = 13.326$ $p < 0.001$): For stimuli with

an interaural correlation change, listeners responded earlier by approximately 80 ms

to 1→ stimuli than to 0→ stimuli, regardless of the step size. Response times were

smaller for larger IAC step sizes. For stimuli with no IAC change (1→1 and 0→0

conditions), there were no differences in the latency of false positive responses,

although the number of false positives was higher in the latter condition (see dotted

bars in Fig. 3.5A).

The behavioral result of greatest interest is the asymmetry in detection rate

and response time between the symmetrical 0→1 and 1→0 conditions. Listeners are

faster and more accurate at detecting a change from correlated to uncorrelated noise

than vice versa.

Interestingly, when asked to describe their experiences of listening to the

changes in interaural correlation, many subjects described the transitions (in both

directions) as *movement*. 0→ transitions were reported as movement towards the

center of the head, whereas 1→ transitions were described as a single focused source

that is "stretching" and moving away from the center towards the two ears. The fact

that the interaural correlation change was perceived as gradual, even though the

physical change was abrupt, may be an indication for the existence of a sliding

binaural temporal integration window (Culling & Summerfield, 1998; Akeroyd &
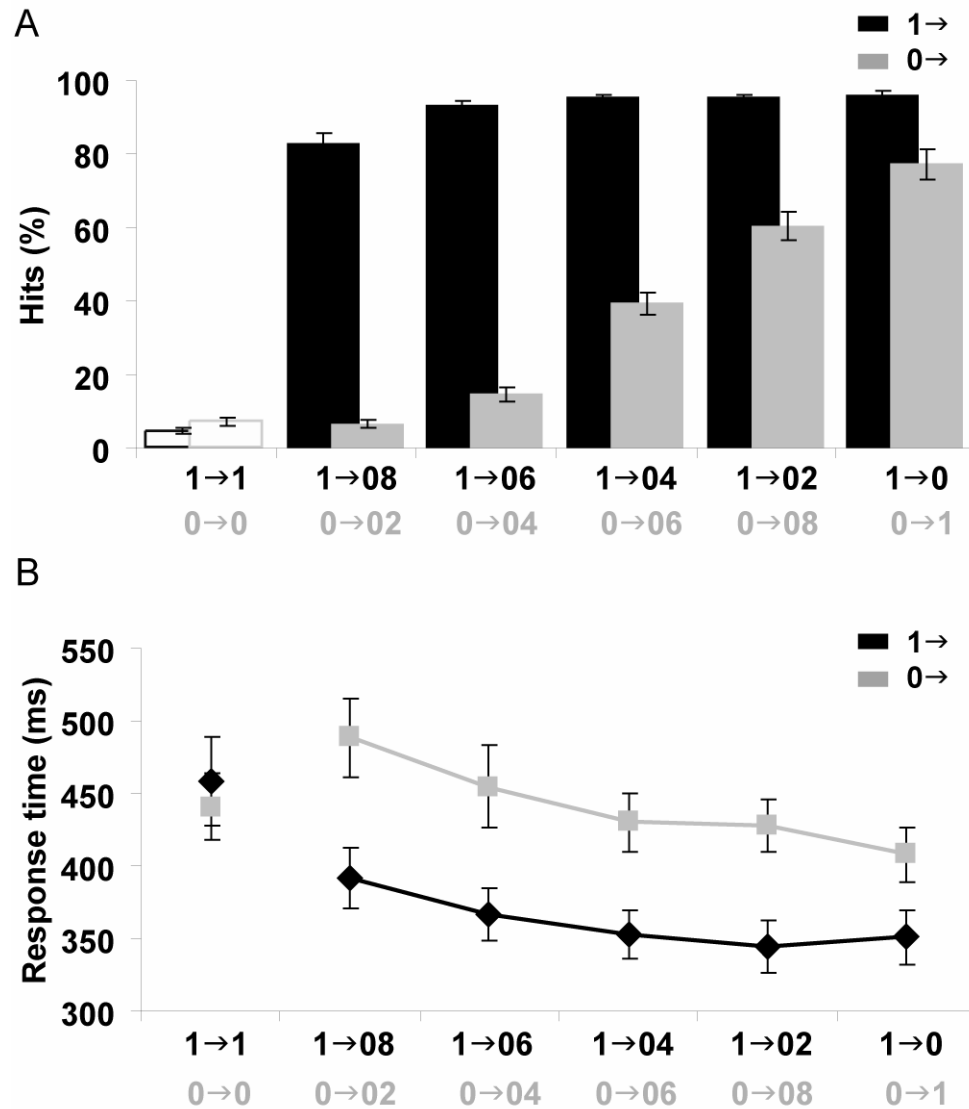
Figure 3.5: Behavioral data: (A) Mean detection rate and (B) mean response time for the different experimental conditions in the behavioral study. Error bars represent 1 std. error. Listeners were almost at ceiling for transitions from IAC = 1 (1→ conditions), but performed much more poorly on same size transitions from IAC = 0   (0→ conditions). The rate of correct IAC change detection increased with larger differences in correlation.

Summerfield, 1999; Boehnke et al., 2002) over which the perceived IAC value is computed. This is discussed further below.

MEG data

Subjects were good at performing the decoy task (modulation detection). The average miss and false positive counts (out of a total of 480 presentations) were 15.1 and 6.5, respectively (std. error 3.01 and 2.53). The average response time was 420.3 ms (std. error 10.79). These behavioral data indicate that subjects were alert and listening to the stimuli, and that task-related attention was focused at the point of change but did not depend on interaural correlation processing.

Waveform and magnetic field distribution analysis reveal that all participants had comparable response trajectories. The auditory evoked responses to the $0\rightarrow1$ condition is shown in Fig. 3.6A. Plotted in grey are the responses for each of the 156 channels averaged over subjects. The root mean square (RMS) over all channels is plotted in red. Responses to other $0\rightarrow$ and $1\rightarrow$ conditions (not shown) are similar to Fig. 3.6A, particularly at the onset. Two aspects of the response are of interest: the peaks following the noise onset, and those following the transition.

The onset response consisted of two peaks, at ~70 ms (M50) and ~170 ms (M150), visible in the grand-averaged data in Fig. 3.6A, both with a spatial distribution characteristic of standard M50 stimulus onset response (Woldorff et al, 1993; Yvert et al, 2001; Chait et al, 2004). Interestingly, the M100 peak, with a spatial distribution opposite that of the M50, that is usually seen at about 100 ms post onset for similar stimuli is greatly reduced here. There appears to be a small

154
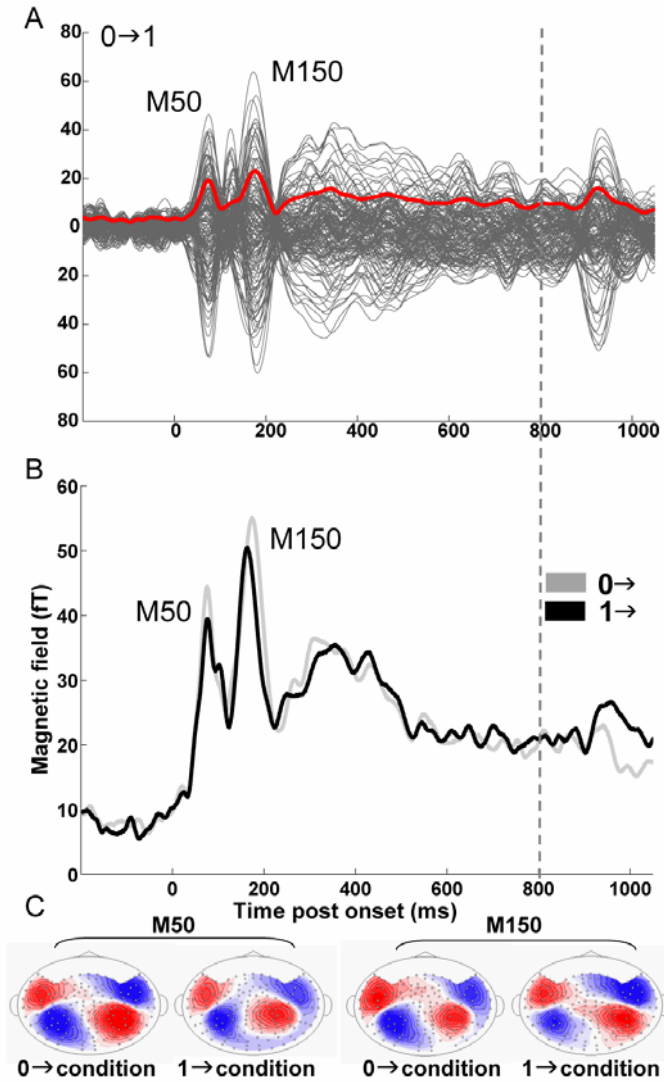
Figure 3.6: MEG data: (A) Example of measured data: Grand-average (average over all subjects for each of the 160 channels; in grey) of the evoked auditory cortical responses to 0→1 stimuli. The root mean square (RMS) over all channels is plotted in red. (B) Group-RMS in the right hemisphere for 0→ (grey) and 1→ (black) stimuli (collapsed across the different IAC step sizes). (C) Contour maps from the grand-average data at the critical time periods (7.5 fT/iso-contour). Source = red, Sink = blue. Onset responses to correlated (1→) and uncorrelated (0→) stimuli were comparable. Both are characterized by a two-peaked noise onset response at approximately 70 ms (M50) and 170 ms (M150) post onset, with similar magnetic field distributions. A notable difference is that M50 and M150 peak amplitudes for each subject were significantly stronger for uncorrelated than correlated noise. The difference in the RMS amplitude between (A) and (B) is a result of (A) being computed from the grand-averaged data whereas (B) are group-RMSs (RMSs of individual subject's RMSs). All statistical analyses were performed on each-hemisphere, subject-by-subject (based on the 20 channels selected for each). The grand-average plot is shown here for illustration purposes only.

155

deflection for some subjects, but in the RMS it is shadowed by the much stronger M50 and M150 responses. This is in contrast to reports by others that describe noise onset responses dominated by a M100 peak (e.g. Soeta et al., 2004). The lack of an M100 is not the effect of channel selection, as the M100 peak is also absent in the RMS over all channels (for example see Fig. 3.6A). Rather, it seems to result from the fact that the subjects' task (detection of modulation in the final portion of decoy stimuli) directed their attention away from the onset. This question has been addressed in another study (Chait et al, 2004). Overall, results suggest that control of the task performed by subjects during recording of brain responses may have a greater importance than is commonly realized.

Onset responses to initial correlated and initial uncorrelated conditions are similar in latency and spatial distribution, but with amplitude stronger for uncorrelated (IAC = 0) than correlated (IAC = 1) noise. Fig. 3.6B shows the group-RMS (RMS of individual-subject RMSs) to 1→ and 0→ conditions (collapsed across the different IAC step sizes) in the right hemisphere. Paired sample t-tests revealed that M50 and M150 peak amplitudes for each subject were significantly stronger for uncorrelated than correlated noise in both hemispheres (df = 17; RH: M50 $t = 2.099$ $p = 0.051$, M150 $t = 2.704$ $p = 0.015$; LH: M50 $t = 2.298$ $p = 0.035$, M150 $t = 3.045$ $p = 0.007$). This finding is perhaps surprising given that amplitudes of onset responses are positively related to loudness (Roberts et al, 2000) and that correlated noise evokes a relatively loud, compact percept whereas uncorrelated noise is perceived as less loud and more diffuse (Blauert & Lindemann, 1986). At the same time, it is in agreement with the Equalization-Cancellation model (Durlach, 1963) that proposes

that the inputs to the two ears are subtracted from each other, and the remainder constitutes the representation of binaural information. Another possible interpretation of this finding is that because the inputs at the two ears do not fuse to a single image, additional neuronal activity is involved in "sorting out" these distributed images. This interpretation is consistent with the shape of EEG binaural interaction components (BIC) of auditory brain stem responses, computed as the difference between the response to binaural stimulation and the sum of the responses to monaural stimulations of the two ears (e.g. Polyakov & Pratt, 1998). BIC are usually negative (the binaural response is smaller than the sum of the monaural responses; see Krumbholz et al, 2005 and references therein) and of greater amplitude for correlated noise than uncorrelated noise (e.g Polyakov& Pratt, 1998), suggesting that activity evoked by binaurally uncorrelated signals undergoes less mutual suppression than activity evoked by correlated signals.

Soeta et al (2004) also found that uncorrelated noise onsets evoked a stronger response than correlated noise onsets. However, in their study stimuli with IAC = 1 and IAC < 1 were alternated, which complicates the interpretation of the results: the weaker responses to stimuli with IAC = 1 may be a result of adaptation and stronger response for stimuli with lower IAC may result from the larger interaural correlation difference with the stimulus that preceded them. In an fMRI study using stimuli with fixed IAC values, Budd et al. (2003) identified a distinct subdivision of lateral HG that exhibited a significant positive relationship between BOLD activity and IAC. Activation differences were larger for IAC near 1 than near 0. The trend is opposite that found in the present study. The apparent discrepancy may result from the current

157

lack of understanding how hemodynamic BOLD responses are related to the electrical physiological brain responses measured by MEG.

Auditory cortical sensitivity to changes in interaural correlation

The transient response due to stimulus onset is followed by a gradual decline to steady-state levels (Fig. 3.7A). The change in interaural correlation at 800 ms produces a response that rides upon this gradual decline, consisting of a prominent peak at approximately 950 ms post onset (150 ms post change). In order to quantify the cortical response to changes in interaural correlation, we subtracted, for each subject and each condition, the time-average amplitude in the 600 ms - 800 ms interval from the time-average amplitude in the 850 ms - 1050 ms interval (Fig. 3.7A). For stimuli for which there was an interaural correlation change, we then subtracted from this statistic its value for the corresponding control condition ($1{\rightarrow}1$ or $0{\rightarrow}0$) for which there was no change in the stimulus. A value significantly different from zero indicates that auditory cortical activity was affected by the interaural correlation change. Fig. 3.7B shows the computed difference for each of the change conditions in the left and right hemispheres. An ANOVA revealed main effects of reference correlation ($F(1) = 10.104$ $p = 0.005$) and size of IAC step ($F(4) = 9.829$ $p < 0.001$): Differences were larger for larger IAC step sizes, and also larger for steps from an initial correlated (IAC = 1) than uncorrelated (IAC = 0) noise. Cortical responses thus parallel ease of detection as measured behaviorally by both accuracy and reaction times. For $1{\rightarrow}$ conditions all differences were significant in both
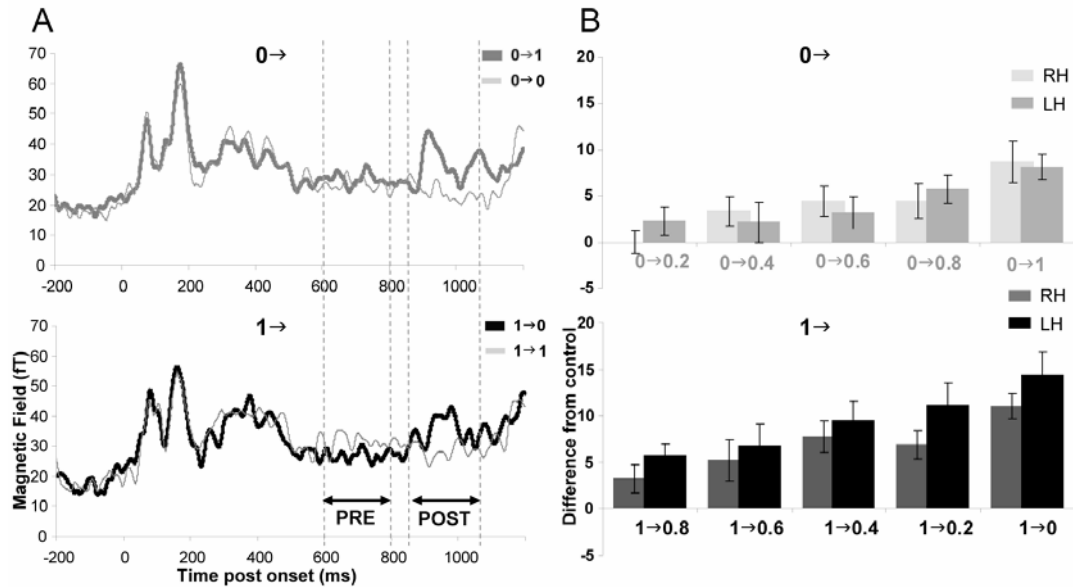
Figure 3.7: Auditory cortical sensitivity to changes in interaural correlation. (A) Schematic example of the procedure used to quantify sensitivity. The figures present, as an example, the group-RMS in the right hemisphere for 0→1 (grey) and 1→0 (black) conditions and their controls (0→0 and 1→1, respectively). To quantify the cortical response to changes in interaural correlation, we subtracted, for each subject and each condition, the time-average amplitude in the 600 ms - 800 ms interval (PRE) from the time-average amplitude in the 850 ms - 1050 ms interval (POST) such that DIFF = POST-PRE. Positive DIFF indicate increase in activity relative to the activity before the change in IAC. For stimuli for which there was an interaural correlation change, we then subtracted from this statistic its value for the corresponding control condition (1→1 or 0→0) for which there was no change. (B) Computed difference for each of the change conditions (0→ conditions in grey; 1→ conditions in black) in the left (dotted pattern) and right (striped pattern) hemispheres. Plotted values are difference between DIFF values in the change conditions vs. control conditions. Positive differences indicate that activity after change in IAC was significantly higher than post 800 ms activity in the control condition. Differences that are significantly different from zero are marked with "*". Error bars represent 1 std. error. These physiological responses resemble behavioral responses in that they are larger for correlated than uncorrelated references, and for larger changes in correlation.

159

hemispheres (planned comparison, $df = 17$; LH: $1\rightarrow0$ $t = 4.465$ $p < 0.001$, $1\rightarrow0.2$ $t = 3.909$ $p = 0.001$, $1\rightarrow0.4$ $t = 3.237$ $p = 0.005$, $1\rightarrow0.6$ $t = 2.229$ $p = 0.04$, $1\rightarrow0.8$ $t = 3.205$ $p = 0.005$; RH: $1\rightarrow0$ $t = 4.858$ $p < 0.001$, $1\rightarrow0.2$ $t = 3.366$ $p = 0.004$, $1\rightarrow0.4$ $t = 3.235$ $p = 0.005$, $1\rightarrow0.6$ $t = 2.773$ $p = 0.013$, $1\rightarrow0.8$ $t = 2.242$ $p = 0.039$). In the case of $0\rightarrow$ conditions, differences were significant for $0\rightarrow1$ and $0\rightarrow0.8$ in the left hemisphere (planned comparison, $df = 17$: $0\rightarrow1$ $t = 4.719$ $p < 0.001$, $0\rightarrow0.8$ $t = 2.539$ $p = 0.021$), and for $0\rightarrow1$, $0\rightarrow0.8$ and $0\rightarrow0.6$ in the right hemisphere ($df = 17$: $0\rightarrow1$ $t = 4.281$ $p < 0.001$, $0\rightarrow0.8$ $t = 2.296$ $p = 0.035$, $0\rightarrow0.6$ $t = 3.102$ $p = 0.006$).

Fig. 3.8 shows the group-RMS of auditory cortical responses to IAC change for $0\rightarrow1$ and $1\rightarrow0$ conditions (other conditions showed a similar response pattern). The change in correlation in the $1\rightarrow$ conditions was characterized by a response with three peaks, around ~70ms (Window 1) , ~130ms (Window 2), and 200ms (Window 3) post change. A deflection is considered a "peak" if it is consistent across subjects (see *methods* section) and has a salient dipolar distribution that is compatible with activity in auditory cortex. The iso-contour magnetic field distribution maps from the grand-average data are also displayed in Fig. 3.8. In contrast, the $0\rightarrow$ condition evoked only one pronounced peak, occurring at a time corresponding to Window 2 (B). Thus, Window 1 contains the first dipolar response to the $1\rightarrow$ transition, whereas the same window shows no coherent response in the $0\rightarrow$ condition. Window 2 shows a prominent peak for $0\rightarrow$ but, remarkably, the dipolar distribution is of opposite polarity from $1\rightarrow$, indicating that activity cannot possibly be resulting from the same neural substrate. Note also that it is of opposite polarity from that in Window 1 for $1\rightarrow$. Thus the later initial response for $0\rightarrow$ does not merely reflect a delayed

activation of the same source. In total these data suggest that the entire sequence of cortical activation involves distinct neural mechanisms in each case: the mechanism that processes transitions from IAC = 1 is different from the mechanism that processes transitions from an IAC = 0. These data are consistent with observations reported in an EEG study by Jones et al. (1991), and their different conclusions are attributable to technological limitations at that time.

The first observed peak for the 1→ conditions (at ~850 ms, Window 1) occurs approximately 80 ms earlier than the first observed peak in the 0→ conditions (at ~930 ms, Window 2). This electrophysiological latency difference may underlie the ~80 ms response time difference observed in our behavioral data. However, the opposite polarities of these "first responses" are a puzzle. One possibility is that behavior is contingent on the activity within distinct neural substrates reflected in Windows 1 (for 1→) and 2 (for 0→). Another possibility is that it is contingent on the same neural substrates, but the activity, visible in Window 1 for 1→ conditions, is either weaker in 0→ conditions, or else delayed and masked by a later activation specific to 0→ conditions (visible in Window 2). Since the data for all stimuli were acquired under identical experimental conditions with the same listeners, any difference in the responses implies differences in processing mechanisms. Results are inconsistent with a general processor that would respond to *any* perceptible change in the steady auditory stimulus conditions (as suggested for example by Jones et al, 1991).
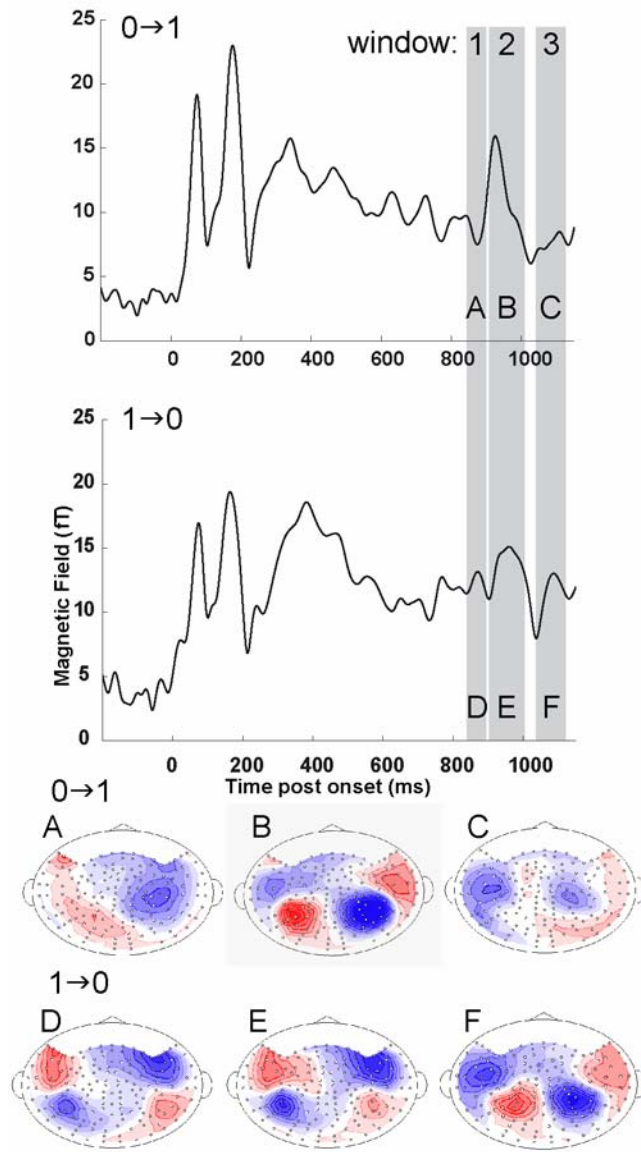
Figure 3.8: RMS (computed over all channels from the grand-average data) of the auditory cortical responses to 0→1 and 1→0 conditions (other conditions show similar response patterns). The response to the change in correlation was characterized by sequential increases in activity in 3 temporal windows ~70 ms (Window 1), ~130 ms (Window 2) and ~200 ms (Window 3) post change in correlation. (A-F) The iso-contour magnetic field distribution maps from the grand-averaged data at the critical time periods (7.5 fT/iso-contour). Source = red, Sink = blue. The responses in the 1→ and 0→ conditions exhibit different magnetic contour map patterns such that 1→ has pronounced dipolar activity in all three time windows (D,E,F), but the 0→ condition has a dipolar pattern only in time window 2 (B).
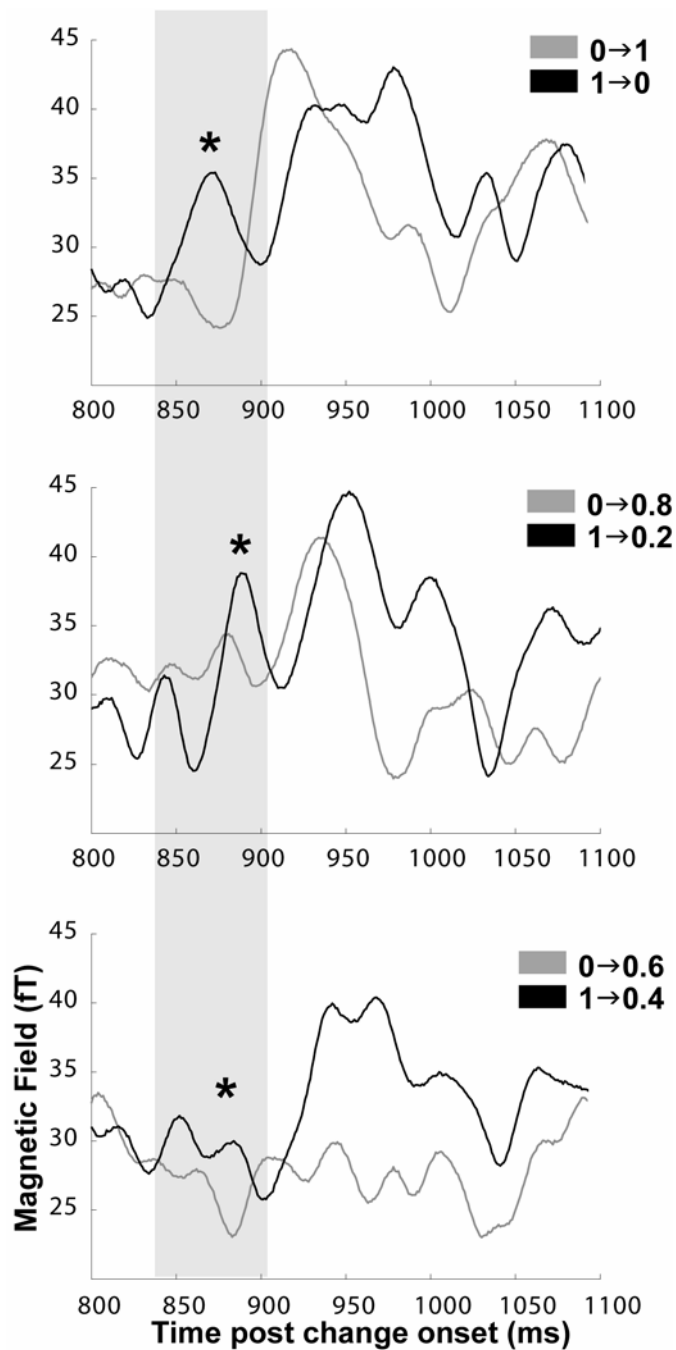
Figure 3.9: Group-RMS in the right hemisphere of the responses to 1→ (black) and 0→ (grey) conditions with an equal IAC change (plotting only those 0→ conditions for which the response to IAC change is significantly different from the control condition). The first increase in activity is evident in all 1→ conditions at approximately 50 ms post change in correlation. Amplitudes in this time window are significantly stronger for 1→ conditions than 0→ conditions with a same-sized change in IAC

In addition to the existence of a coherent dipolar pattern in Window 1, the 1→ conditions always had higher amplitude in that window relative to the corresponding 0→ conditions. This effect is shown in Fig. 3.9 for the conditions for which 0→ activity is significantly different from its control (see Fig. 3.7B; 1→0/0→1, 1→0.2/0→0.8, 1→0.4/0→0.6). Significance was assessed with the Bootstrap method (500 iterations; balanced; Efron & Tibshirani, 1993), a computationally intensive re-sampling method that allows the treatment of situations in which the exact sampling distribution of the statistic of interest is unknown. For each subject, the RMS of the 0→ condition was subtracted from the applicable (same IAC distance) 1→ condition and the difference vectors were bootstrapped. We computed the distribution of bootstrap amplitudes at the peak of the mean difference vector in window 1 for each of the three condition pairs (RH: 871 ms, 891 ms and 883 ms; LH 887 ms, 874 ms and 880 ms) and counted the percent of iterations for which the amplitude difference was smaller or equal to zero (*perct*). A value of *perct* that is lower than the *a priori* set 5% level was considered to indicate a significantly higher amplitude in the 1→ conditions relative to the corresponding 0→ conditions (RH: 0→1/1→0 *perct* = 2.6%, 0→0.8/1→0.2 *perct* = 4.6%, 0→0.6/1→0.4 *perct* = 1.4%; LH: 0→1/1→0 *perct* = 1.2%, 0→0.8/1→0.2 *perct* = 4.8%, 0→0.6/1→0.4 *perct* = 2.8%).

3.1.3 Discussion

We used behavioral methods and whole-head MEG recording to measure responses to the same binaural wide-band noise stimuli. For a given step size in interaural correlation, subjects detected transitions from IAC = 1 more accurately and rapidly

than from IAC = 0. This is consistent with previous studies (Pollack & Trittipoe, 1959a,b; Gabriel & Colburn, 1981; Culling et al, 2001) reporting that equal steps in IAC are not equally salient perceptually in the vicinity of IAC=0 and IAC=1. However, our results go further by showing an effect of the *sign* of IAC change, most clearly obvious for the symmetric 1→0 and 0→1 stimuli. This suggests that IAC discriminability might not be adequately described by distance along an internal decision axis, as a distance is by definition symmetric. A similar asymmetry is prominent in brain responses.

Our behavioral task required subjects to detect a change in interaural correlation, whereas the cortical responses were passive responses to IAC change. Nevertheless, the relationship between the strength of the measured cortical responses to the different conditions (Fig. 3.7B) paralleled behavioral performance (Fig. 3.5A): brain responses were more sensitive to transitions from IAC = 1 (1→ conditions) than to transitions from IAC = 0 (0→ conditions). Sensitivity in all cases increased with IAC difference. In this respect our behavioral and brain studies are consistent with each other and with prior literature. In addition, the first salient response to steps from IAC = 1 occurred earlier than from IAC = 0, which parallels the latencies measured behaviorally and may conceivably be the neural correlate of the observed behavioral response-time differences. Overall, the earliest observed cortical responses already reflected the asymmetry seen in behavior. What is new is the conclusion, derived from the different polarities of the magnetic field distribution, that 1→ and 0→ transitions evoke activity within different cortical circuits. This result is unexpected, as one would assume all aspects of IAC processing (and changes thereof) to engage

common binaural processing mechanisms, and it may shed light on the nature of the computing involved. This finding should be replicable in fMRI (previous studies used fixed IAC values) as well as physiology.

The $1\rightarrow$ and $0\rightarrow$ conditions differ in both the direction of IAC change (increase or decrease), and in the value of initial correlation (1 or 0). The present study cannot determine whether the observed differential processing is related to the reference correlation or the direction of correlation change. This issue can be resolved in future experiments by studying stimuli with initial correlations different from 0 or 1 (such as $0.5\rightarrow 0$ vs $0.5\rightarrow 1$ or $0.8\rightarrow 1$ vs. $1\rightarrow 0.8$).

It is unclear at which level occurs the split into distinct processing streams, or the introduction of the 80 ms latency difference between $1\rightarrow$ and $0\rightarrow$ conditions. The MEG responses we record originate from auditory cortex. The computation of interaural correlation is thought to begin at the MSO, where information from the two ears converges on coincidence detectors that perform a form of interaural cross-correlation (Jeffress, 1948; Yin & Chan, 1990; Carr, 1993; Joris et al., 1998). From there, the binaural information pathway projects to the IC, medial geniculate body and cortex. Animal electrophysiological recordings at MSO are rare, but recordings in the IC show correlates of binaural unmasking (Jiang et al., 1997ab; Palmer et al., 2000) and responses that are influenced by the interaural correlation of stimuli (Palmer et al, 1999). The question arises as to whether sensitivity to binaural coherence is determined by processes at IC (in the same way that basic masking is determined by processes in the auditory nerve), and relayed from there, or if later stages are involved in measuring interaural correlation.

An aspect of interaural correlation processing that has been hypothesized to involve cortical mechanisms and may be related to processes observed here is "binaural sluggishness": it has been demonstrated that human listeners become less sensitive to time-varying changes in interaural correlation as the change rate is increased (Grantham, 1982). This suggests that listeners compute the effective IAC value over a binaural integration window that in turn influences detection in binaural unmasking situations (Grantham & Wightman, 1979; Culling & Summerfield, 1998; Akeroyd & Summerfield, 1999). Joris et al. (2006) did not find correlates of sluggishness in IC: single units followed modulations of IAC at rates an order of magnitude higher than the behavioral threshold, suggesting that the site of temporal integration is higher upstream, possibly in cortex.

Binaural sluggishness may be functionally justified by the need to acquire binaural information over a time sufficient to eliminate random fluctuations. These temporal integration mechanisms may underlie the cortical processing speed difference observed here. As discussed above, Joris et al (2006) showed that IC neurons react promptly to IAC changes. The longer time it takes cortical mechanisms to respond to one condition vs. another can be explained in terms of a central system that integrates the instantaneous information received from IC over time until it has reached a sufficient level of reliability (Shinn-Cunningham & Kawakyu, 2003). The amount of temporal integration may be constant or vary depending on the stimulus and/or task. For example, supposing that activity over a population of neurons within the MSO is accurately represented by the stimulus cross-correlograms in Fig. 3.3, a mechanism that scans this activity would be able to respond relatively soon after a

change from a reference correlation of 1 (Fig. 3.3A) because such a change "destroys" the orderly arrangement of ridges and valleys that characterizes the response to correlated noise. Conversely, the opposite change ($0{\rightarrow}1$) would take longer to detect because uncorrelated noise is already characterized by random changes in the activation across the neural array, so it would require more time to determine that the sudden order in the stimulus is not merely a random fluctuation. A similar account can also be provided in terms of the equalization-cancellation model (Durlach, 1963): if binaural information is represented as subtraction from the two ears then binaural noise with IAC = 1 would be represented as a 800 ms-long "zero" whereas noise with IAC = 0 would be represented as 800 ms-long activation with high variability. For $1{\rightarrow}0$ stimuli, the change after 800 ms would be evident as a sudden change from 0 to a positive number. For $0{\rightarrow}1$, the change at 800 ms is preceded by random fluctuations and the system would need to wait longer to detect the change.

Response latencies measured in auditory cortex provide an *upper-limit* for the size of the binaural integration window: about 50 ms for transitions from correlated noise and about 130 ms for transitions from uncorrelated noise. These estimates are similar to those derived from behavioral measurements (Culling & Summerfield, 1998; Akeroyd & Summerfield, 1999; Boehnke et al., 2002). The different integration times required for $1{\rightarrow}$ and $0{\rightarrow}$ transitions might conceivably be implemented by a single mechanism with a variable integration time. Another possibility is that the process passes via two successive integration mechanisms: an initial obligatory integration window and a subsequent integration window, provided by a separate

168

neural substrate, that is only required in the 0→ condition in order to reach a sufficient level of certainty that there has been a change. Such a model would explain the activation of *distinct* neural populations for the two kinds of transitions.

That it takes longer to react to changes from a disordered state to an ordered state than vice versa may be a general attribute of perceptual phenomena. For example, data on the perception of dynamic random-dot stereograms (Julesz and Tyler, 1976) are very similar to those obtained in the auditory domain with noise signals. The visual stimuli strongly parallel ours, consisting of frame sequences in which the left and right frames are either identical (interoccular correlation IOC = 1) or uncorrelated (IOC = 0). Subjects' ability to detect changes in IOC from 0 to 1 and vice versa revealed an asymmetry, similar to the results presented here. Julesz and Tyler (1976) liken this effect to the physical concept of entropy: perceptual phenomena require more effort/time to build up representations (go to an orderly state) than to destroy them (go to a less ordered state). This account also offers an additional interpretation to the observed three-staged processing of changes in interaural correlation. The first peak (Window 1), only visible in the 1→ conditions may reflect the destruction of (or detection of a violation in) the representation of the correlated noise, while the second peak (Window 2), visible in both conditions, but having different source properties may underlie the construction (or attempts at the construction) of a new perceptual order.

The strong similarities in detecting changes in correlation between vision and audition may indicate that the statistical rules that determine the size of the integration windows are modality-independent and are not special to a particular

neural substrate. This observation, together with the findings reported here, may provide a basis for further examinations of how the central nervous system computes and represents changes in the environment.

3.2 Human auditory cortical processing of transitions between order and disorder

Auditory environments are characterized by constant change, occurring over a continuum of time scales. Sound, by its essence as a mechanical wave, changes from moment to moment. Changes on a larger scale occur within the sound emitted by a source, such as the formant transitions during the utterance of a sentence, a sequence of notes in a musical tune, or the periodic hum of an engine. On a yet larger scale, change is introduced in the ongoing waveform when sound generating objects appear or disappear.

Much of auditory experimentation uses signals that are presented between quiet inter-stimulus intervals, but this is rarely the case in real environments. To detect the appearance of a new auditory object, a listener must acquire the statistics of the ongoing sound, compare the real time incoming information to this internal model, and react if a deviation occurs. In order to do this successfully, one has to be able to differentiate the sort of change that is expected from an auditory object, from abnormal deviations that indicate a new sound source distribution. This implies that listeners should be very sensitive to the statistical distributions of sounds in their surroundings. There is evidence that listeners do indeed build up representations of the statistics of their auditory environments, and that they do so pre-attentively, without conscious effort and in situations where this information is not immediately required for behavior. For example, it has been demonstrated that adults and babies automatically track the statistical properties of sound combinations and are able, with

very brief exposure, to detect and use information about the probability with which sounds co-occur to infer word boundaries (Saffran et al, 1996; Saffran, 2001; Maye et al., 2002) or to appreciate regularity in music (Saffran et al., 1999; Tillmann & McAdams, 2004).

Recent electrophysiological studies in animals demonstrate neural adaptation to dynamic changes in stimulus statistics, serving to adjust the performance of the auditory system to the instantaneous properties of the acoustic environment (Ulanovsky et al, 2003; Ulanovsky et al, 2004; Dean et al, 2005). Change detection in humans is usually studied with the Mismatch Negativity (MMN) paradigm: Rare deviant stimuli occasionally replacing repeated, 'standard' stimuli elicit a brain potential, peaking at about 200ms post change onset. (Kujala & Naatanen, 2003). The MMN is thought to reflect a discrepancy between the input from a deviant stimulus and the memory trace or expectations generated by the repeated standard stimulus (Naatanen, 1985; Naatanen, 1992; Sams et al., 1993). It has been suggested that increased negativity, elicited by deviant stimuli, is due to processes related to updating the internal representation when a previously registered regularity is violated (Winkler et al., 1996; Winkler, 2003). MMN experiments are usually conducted with silent intervals between relevant stimuli, so that the time at which a stimulus is compared to the preceding one is defined by the experimenter. The experiments to be described here target a stage that directly precedes the stage probed with standard MMN techniques. Faced with an ongoing continuous stimulation, how does the system *know* that regularity has indeed been violated? Or that a new regularity has emerged?

172

In order to study the dynamics of change detection, we employ simple signals that embody certain characteristics of natural auditory objects. The current study uses pure tones that are either constant ('C'; a single tone for the entire duration of the stimulus), randomly varying ('R'; a random sequence of tone pips) or changing from constant to random or vice versa ('CR' and 'RC', respectively; Figure 3.10). Constant tones and random sequences of tone pips are intended to model different states of regularity vs. 'randomness', as they might occur in the environment. By recording brain activity while subjects listen to these sounds, we can measure at which point in time the different transitions are detected, and characteristics of the response give us insights as to the mechanisms. Such transitions are relevant to auditory scene analysis because the appearance of an object within a noisy background may be manifest as a sudden emergence of order within the ongoing random signal. Likewise, the appearance of an unexpected event can be understood as a disruption of a previous, predictable, order.

From a theoretical point of view, one can argue that an ideal observer is able to immediately detect the transition in the 'constant-to-random' case. The first waveform sample that violates the regularity rule suffices to signal the transition. The opposite transition – from random to constant - takes longer to detect, because the observer must wait long enough to distinguish the transition from a momentary 'lull' in the fluctuation, such as might occur by chance. The amount of time an optimal listener would have to wait in order to detect the change depends on the statistical properties of the stimulus (e.g. DeWeese and Zador, 1998). In the case of our stimuli the relevant parameter is pip duration. The ideal observer must wait *at least* the
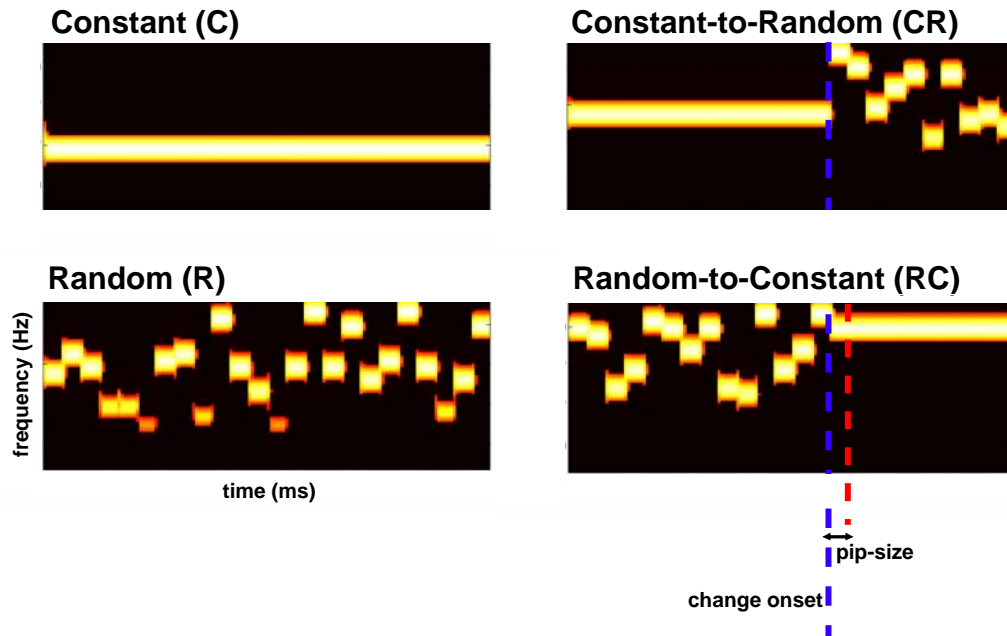
Figure 3.10: Examples of the 4 stimulus configurations used in the experiment. Signals were either a constant tone (C)random sequence of tones (R) or contained a transition from constant to random (CR) or vice versa (RC). The blue dotted line marks the onset of change. An ideal observer is able to detect the CR transition immediately, however the RC stimuli are only distinguishable from the R (control) condition after one pip-size has elapsed (red dotted line).

The plots represent 'auditory' spectrograms, generated with a filterbank of 1/ERB wide channels (Equivalent Rectangular Bandwidth, Moore & Glasberg, 1983) equally spaced on a scale of ERB-rate. Channels are smoothed to obtain a temporal resolution similar to the Equivalent Rectangular Duration measured by Plack and Moore (1990).

duration of a pip in order to detect the change from random to constant (see Figure 3.10).We used pip sizes of 15 ms, 30 ms and 60 ms (and in a control condition, 120 ms). If the change detection system is indeed adjustable and accurate, detection of change should depend on pip-size duration and occur faster for short pip stimuli than for stimuli consisting of longer tone-pips. We are specifically interested in preattentive mechanisms by which the auditory system automatically parses incoming information regardless of the task at hand. In our paradigm, listeners who are naïve to the purposes of the study are presented with these tonal signals while performing a decoy task (detection of noise) unrelated to change processing. The task serves merely to keep subjects awake and attentive to the auditory modality. By analyzing the responses to the transitions in the tonal stimuli interspersed between the decoy targets, we can determine whether the auditory system adjusts to the statistical properties of ongoing sounds, whether it does so automatically (preattentively), and with what accuracy.

3.2.1 Materials and Methods

Subjects

Twenty-four subjects (mean age 20.2 years, 14 female), took part in the experiments. One subject was excluded from the analysis of the behavioral data because of inability to perform the change-detection task (over 50% misses). All participants were right handed (Oldfield, 1971), reported normal hearing, and had no history of neurological disorder. The experimental procedures were approved by the University of Maryland institutional review board and written informed consent was obtained

from each participant. Subjects were paid for their participation.

Stimuli

MEG study: The signals were 1440 ms in duration, consisting of an initial 840 ms pre-transition segment (either random or constant, as defined above), immediately followed by a 600 ms post-transition segment (either random or constant). Controls were 1440ms random or constant signals, resulting in a total of 4 configurations (Figure 3.10).

We generated 40 signals for each of the 4 patterns (C, R, CR, RC). Frequencies were drawn from 20 frequency values equally spaced on a log scale between 222-2000 Hz. The frequency of each tone and pip was randomly chosen from this set, with the following two constraints: The step at the transition between segments (at 840ms post onset) had to be at least 20% in order to make it perceptually salient. Each frequency had to occur the same number of times (twice) at each time point, to allow summation of MEG responses over epochs (MEG responses are known to be frequency-dependent and an unbalance in frequency content might have created artifactual differences between conditions). For symmetry, CR and RC stimuli were created as mirror images and then trimmed to the required duration. Within each random segment, the probability that two consecutive pips shared the same frequency was about 5%.

We used 3 pip durations (15, 30 and 60 ms) in the main experiments, and 120 ms in a control. Pip durations were presented by blocks. Tone and pip onsets and offsets were ramped with 3ms cosine-squared ramps. In each of the 3 blocks, subjects heard 120 repetitions of every one of the 4 patterns. To counter-balance potential

order effects, the order of blocks was randomized between subjects (Latin-square design). Within each block the order of presentation was randomized, with the inter-stimulus interval (ISI) randomized between 600-1400 ms.

In addition to these tonal stimuli, for which MEG responses were recorded, the stimulus set included a proportion (33%, or 240 per block) of 200 ms wide-band noise bursts (ramped on and off with 10ms cosine-squared ramps). Subjects were instructed to detect these "decoy" stimuli. To do so, listeners had to remain alert and attend to the auditory modality, but the task did not require any processing of the tonal changes that were the focus of our study.

The stimuli were created off-line and saved in 16-bit stereo wave format at a sampling rate of 44 kHz. The signals were delivered to the subjects' ears with a tubephone (E-A-RTONE 3A 50 ohm, Etymotic Research, Inc) attached to E-A-RLINK foam plugs inserted into the ear-canal and presented at a comfortable listening level.

Behavioral study: The stimuli for the behavioral study were identical to the MEG stimuli except that the noise bursts were not included. As for MEG, each listener was presented with 3 blocks (one for each pip size) of 4 patterns (x 120 presentations) in random order. For each listener the order of blocks was the same as for MEG, but the order of stimuli within a block was randomized anew. ISIs were as for MEG. Subjects were instructed to press a mouse button as fast as they could when they heard a transition (random to constant or vice-versa) in the ongoing sound.

Procedure

The MEG recording was always conducted before the behavioral session (with an average of about one week between the two sessions). This was important in order to make sure that during the MEG session subjects were naïve as to the real purpose of the investigation and did not focus special attention on tonal transitions.

MEG study: The subjects lay supine inside a magnetically shielded room. In a pre-experiment, ran just before the main experiment, subjects listened to 200 repetitions of a 1 kHz 50 ms sinusoidal tone (ISI randomized between 750-1550 ms). Responses were used to verify that signals from auditory cortex had a satisfactory signal to noise ratio (SNR), that the subject was positioned properly in the machine, and to determine which MEG channels best respond to activity within auditory cortex. In the experiment proper (about 1.5 hours), subjects listened to stimuli while performing the noise burst detection task as described above. They were instructed to respond by pressing a button, held in the right hand, as soon as a noise burst appeared. The instructions encouraged speed and accuracy. Stimulus presentation was divided into runs of 160 stimuli. Between runs subjects were allowed a short rest but were required to stay still.

Behavioral study: The experiment lasted about 1 hour. Subjects sat in a quiet darkened room and were instructed to press a mouse button held in their right hand as soon as they detect a change in the reference sound. No feedback was provided. Responses and response times were stored and analyzed. The experiment was divided into runs of 200 stimuli. Between runs subjects were allowed a short rest but prohibited from getting up or removing the ear pieces. Before the experiment proper, subjects completed a short practice run with feedback. The stimulus delivery

hardware, software, and headphones were identical to those used in the MEG recording.

Neuromagnetic recording and data analysis

Methods and analysis are described in more detail in Chait et al (2005). The magnetic signals were recorded using a 160-channel, whole-head axial gradiometer system (KIT, Kanazawa, Japan). Data for the pre-experiment were acquired with a sampling rate of 1kHz, filtered online between 1 Hz (hardware filter) and 58.8 Hz (17 ms moving average filter), stored in 500 ms stimulus-related epochs starting 100 ms pre-onset, and baseline-corrected to the 100 ms pre-onset interval. Data for the main (change detection) experiment were acquired continuously with a sampling rate of 0.5 kHz, filtered in hardware between 1 and 200 Hz with a notch at 60 Hz (to remove line noise), and stored for later analysis. Effects of environmental magnetic fields were reduced based on several sensors distant from the head using the CALM algorithm (Adachi et al., 2001), and responses were then smoothed by convolution with a 39 ms Hanning window (cutoff 55 Hz). These are standard signal processing methods; further processing is described below.

Auditory evoked responses to the pure tones presented in the pre-experiment were examined, and the M100 onset response identified. The M100 is a prominent and robust (across listeners and stimuli) deflection at about 100 ms post onset, and has been the most investigated auditory MEG response (see Roberts et al, 2000 for review). It was identified for each subject as a dipole-like pattern (i.e. a source/sink pair) in the magnetic field contour plots distributed over the temporal region of each hemisphere. In previous studies, under the same conditions, the resulting M100

179

current source localized to the upper banks of the superior temporal gyrus in both hemispheres (Hari, 1990; Pantev et al, 1995; Lütkenhöner & Steinsträter, 1998). For each subject, the 20 strongest channels at the peak of the M100 (5 in each sink and source, yielding 10 in each hemisphere) were considered to best reflect activity in the auditory cortex and thus chosen for the analysis of the experimental data.

Stimulus-evoked magnetic fields, measured outside the head by MEG, are generated by synchronous neuronal currents flowing in tens of thousands of cortical pyramidal cells on the supratemporal gyrus (Hämäläinen et al., 1993). This electromagnetic fluctuation is detected as a magnetic dipole characterized by position, orientation and strength. Because of the location of the source inside a cortical fold, responses from auditory cortex typically manifest a characteristic dipolar distribution (source/sink pairs that are anti-symmetric across the two hemispheres). In the following figures we use 2-D contour maps to display this information.

In this study we investigate the temporal characteristics of the brain responses evoked by our stimuli. Two measures of dynamics of cortical processing are reported: the amplitude time course (increases and decreases in activation) as reflected in the root mean square (RMS) of the selected channels, and the accompanying spatial distributions of the magnetic field (contour plots) at certain times post onset. For illustration purposes, we plot the group-RMS (RMS of individual RMSs, computed on the basis of the channels chosen for each subject) or the grand-average (average over all subjects for each of the 160 channels) but the analysis is always performed on a subject-by-subject, hemisphere by hemisphere, basis, using the RMS values of 10 channels chosen for each subject in each hemisphere.

In the main experiment, 1600 ms epochs (including 200 ms pre onset) were created for each of the twelve stimulus conditions (3 pip-sizes × 4 patterns). Epochs with amplitudes larger than 3 pT (~5%) were considered artifactual and discarded. The rest were averaged, low-pass filtered at 30 Hz (67 point wide Hanning window), and base-line corrected to the pre-onset interval. In each hemisphere, the root mean square (RMS) of the field strength across the 10 channels, selected in the pre-experiment, was calculated for each sample point. Twenty-four RMS time series, one for each condition in each hemisphere, were thus created for each subject.

To evaluate congruity across subjects, the individual RMS time series were combined into twenty-four group-RMS (RMS of individual RMSs) time series. Consistency of peaks in each group-RMS was automatically assessed with the Bootstrap method (500 iterations; balanced; Efron & Tibshirani, 1993). The consistency, across subjects, of magnetic field distributions at those peaks was assessed automatically by dividing the 20 channels chosen for each subject into 4 sets (5 channels each): left temporo-frontal, left posterior-temporal, right temporo-frontal, right posterior-temporal (see Fig. 3.11). For each set, the activation was averaged over a 20ms window defined around the group-RMS peak, and the set was classified as either a "sink" (negative average amplitude) or a "source" (positive average amplitude). If the majority of subjects showed the same sink-source configuration, the pattern was considered consistent across subjects.
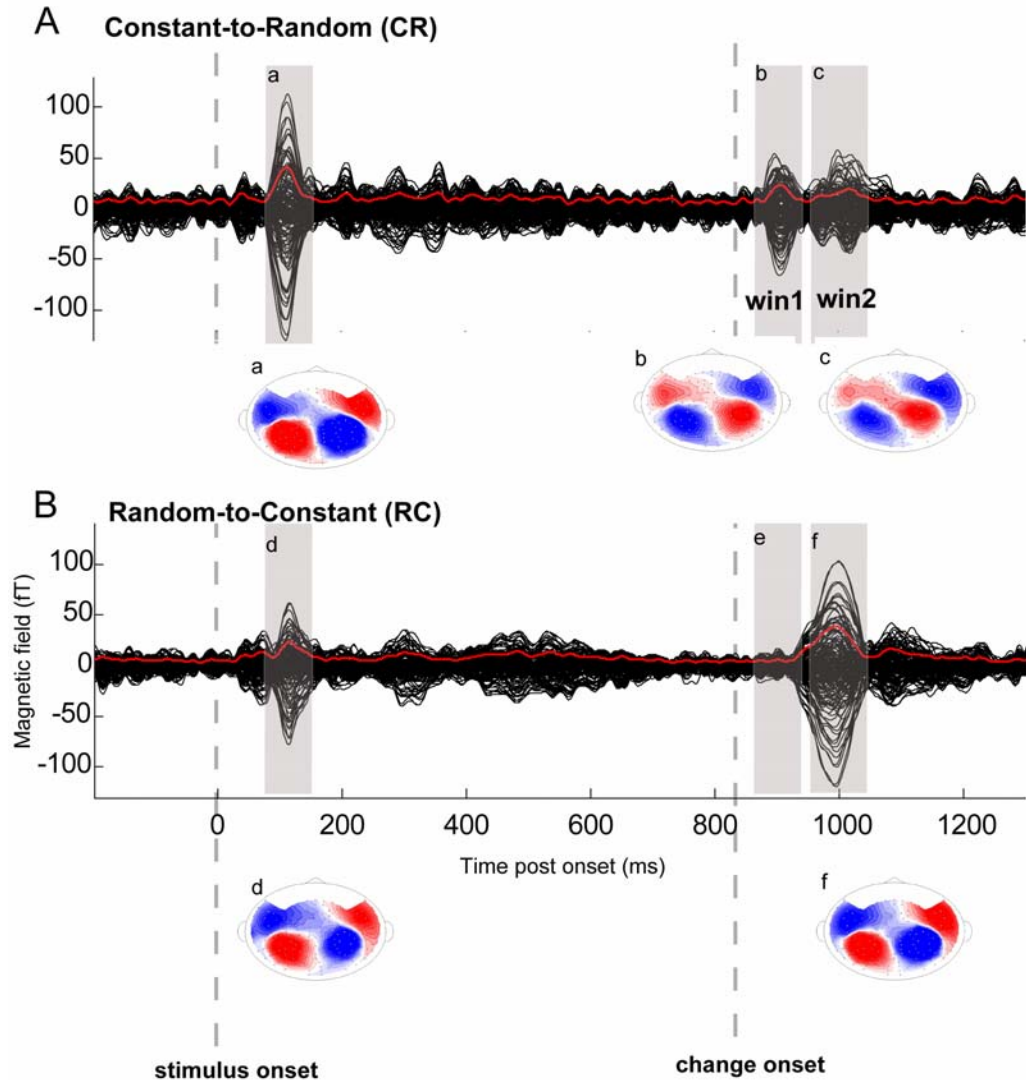
Figure 3.11: Example of measured data. Grand-average (average over all subjects for each of the 156 channels; in black) of the evoked auditory cortical responses to CR (A) and RC (B) stimuli. The root mean square (RMS) over all channels is plotted in red. Contour maps at the critical time periods are also provided (7.5 fT/iso-contour). Source = red, Sink = blue. Onset responses to CR and RC stimuli were comparable. Both are characterized by a pronounced M100 onset response at approximately 110 ms post onset, with similar magnetic field distributions ('a' and 'd', respectively). The response to the transition was characterized by sequential increases in activity in 2 temporal windows ~70 ms (Win 1), ~150 ms (Win 2) post change onset. CR conditions exhibited peaks in both time windows, whereas RC conditions were characterized by a peak in win2 only. Notably, the magnetic field pattern of the responses also differed. Both peaks in the CR stimuli exhibited an M50 like dipolar pattern ('b' and 'c') whereas the contour map corresponding to the RC response was characterized by an opposite dipolar distribution ('f').

3.2.2 Results:

MEG experiment:

Decoy Task

Subjects were accurate and fast at performing the decoy task (detecting the noise bursts). The average miss and false-positive counts were 4.5 and 1.5, respectively (out of a total of 240 presentations per block). The average response time was 375 ms. None of the measures differed between blocks. This implies that subjects were alert and listening to the stimuli throughout the experiment.

MEG responses

Magnetic waveform and field distribution analysis reveal that all participants had comparable response trajectories. The auditory evoked responses to constant-to-random and random-to-constant transitions are shown in figure 3.11, for a pip size of 15 ms. The origin of the time scale coincides with the onset of the signals. The change is introduced at 840ms post onset. Plotted in black are the data for each of the 157 channels, averaged over subjects. The root mean square (RMS) over all channels is plotted in red. The MEG activity evoked by the stimuli exhibits an onset response, about 100ms after the onset of the stimuli and a later response related to processing the change in the stimuli, which begins at about 900ms post onset (60ms post change). These two aspects of the response are discussed, in turn, below.
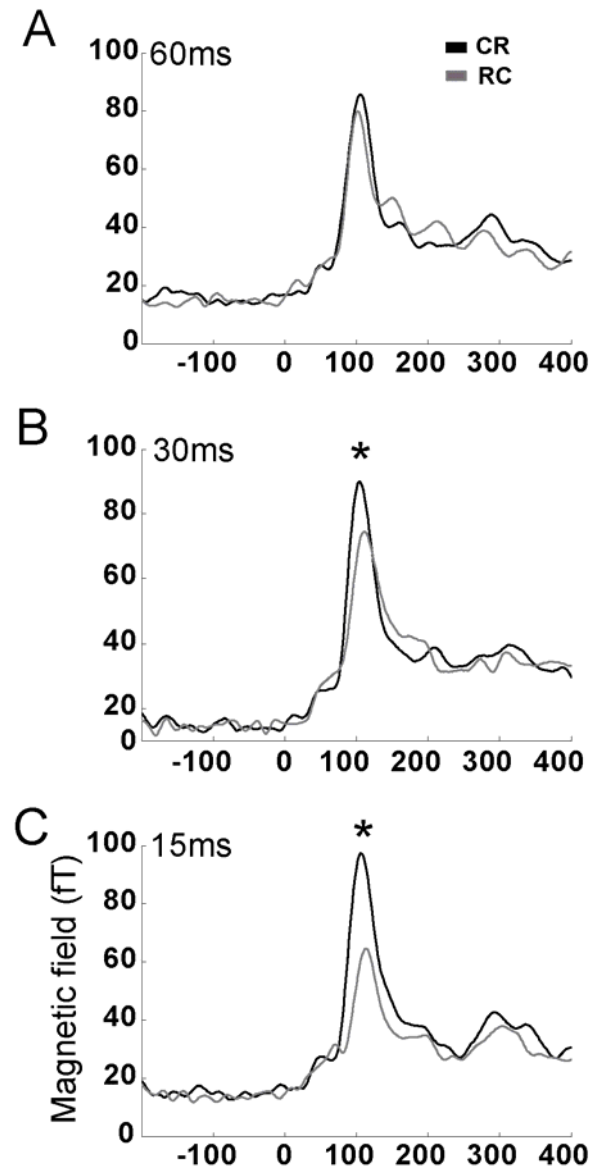
Figure 3.12: Group-RMS in the left hemisphere for CR (black) and RC (grey) stimuli for each of the pip-size conditions. Asterisks denote significant amplitude differences. Apart from an amplitude difference, onset responses to all CR and RC conditions were characterized by identical temporal dynamics. All conditions produced a prominent M100 onset response with a characteristic field distribution (illustrated in figure 3.11). The difference in the RMS amplitude between the plots here and those in figure 3.11 is a result of figure 3.11 being based on the grand-averaged data whereas here we plot group-RMSs (RMSs of individual subject's RMSs). All statistical analyses were performed on each-hemisphere, subject-by-subject (based on the 20 channels selected for each). The grand-average plot is shown here for illustration purposes only.

Onset Responses: Figure 3.12 shows a comparison of onset responses to CR (in black) and RC (in grey) stimuli for each pip-size condition. Onsets of all conditions had similar dynamics (latency and shape of the deflection) and magnetic field distributions. All conditions produced a prominent M100 onset response at approx 110 ms post onset with a characteristic field distribution (see Figure 3.12 'a' and 'd'). There were however some minor amplitude differences. Onset M100 amplitude was greater for CR than RC patterns at 30ms (Fig 3.12B; repeated measures t-test, corrected df=23, LH: t=3.51, *p=0.002;* RH: t=2.46 *p=0.021*) and 15 ms (Fig 3.12C; LH: t=4.6, *p<0.001;* RH: t=6.7 *p<0.001*), but not 60 ms (Fig 3.12A). Amplitudes did not differ between pip-size conditions for CR conditions (black lines in Fig3.12 A, B and C), but increased progressively with pip size for RC conditions (grey lines in Fig3.12 A, B and C). The progressive increase in amplitude difference between corresponding constant and random M100 onset responses as pip size decreased is consistent with previous reports of a ~40ms temporal window of integration during which stimulus attributes are accumulated in processes leading up to the formation of the M100 peak (Gage & Roberts 2000; Roberts et al, 2000; Gage et al, 2006).

Transition response: Unlike onset responses which, as discussed above, are very similar across conditions, transition responses are distinctly disparate (Fig. 3.11). Responses to CR and RC transitions exhibit markedly different temporal dynamics and field distributions. The change from a constant tone to a random sequence of tone pips evokes two consecutive deflections, at about 70 and 150 ms post change onset ('win 1'and 'win 2' in Fig 3.11A respectively). The opposite transition evokes only

one peak, occurring about 150ms post change onset ('win2' in Fig 3.11B). The responses are also characterized by different magnetic field distributions. Remarkably, the dipolar distribution of the RC transition response peak in Figure 3.11B ('f') is of *opposite polarity* from the CR transition response peaks in Figure 3.11A ('b' and 'c') indicating that activity results from a different neural substrate. For that reason, we can conclude that the later response to the RC transition does not merely reflect delayed activation compared to the CR transition. Rather, the data suggest that the sequence of cortical activation is distinct in each case: processing of transitions in each direction involve different neural populations. These findings parallel those reported in Chait et al (2006) for changes in interaural correlation, if one equates 'constant tone' to 'interaurally correlated noise' and 'random sequence' to 'interaurally uncorrelated noise': the same asymmetry was observed between responses to correlated-to-uncorrelated vs uncorrelated-to-correlated transitions. This is further discussed below.

Figure 3.13A illustrates the response to change in the right hemisphere (left hemisphere responses exhibit the same properties). The transition from constant to random (top) is characterized by two successive deflections at around 70 ms (win1) and 150 ms (win2) post change. The amplitudes and latencies of these responses do not differ significantly between windows and between pip sizes (Fig 3.13B). The transition from random to constant is characterized by only one deflection, within win2, with an amplitude and latency that varies with pip size. Figure 3.13C shows a comparison of RC response amplitudes (dark colors) with the temporally corresponding (win2) peak amplitudes in the CR conditions (light colors). Unlike in
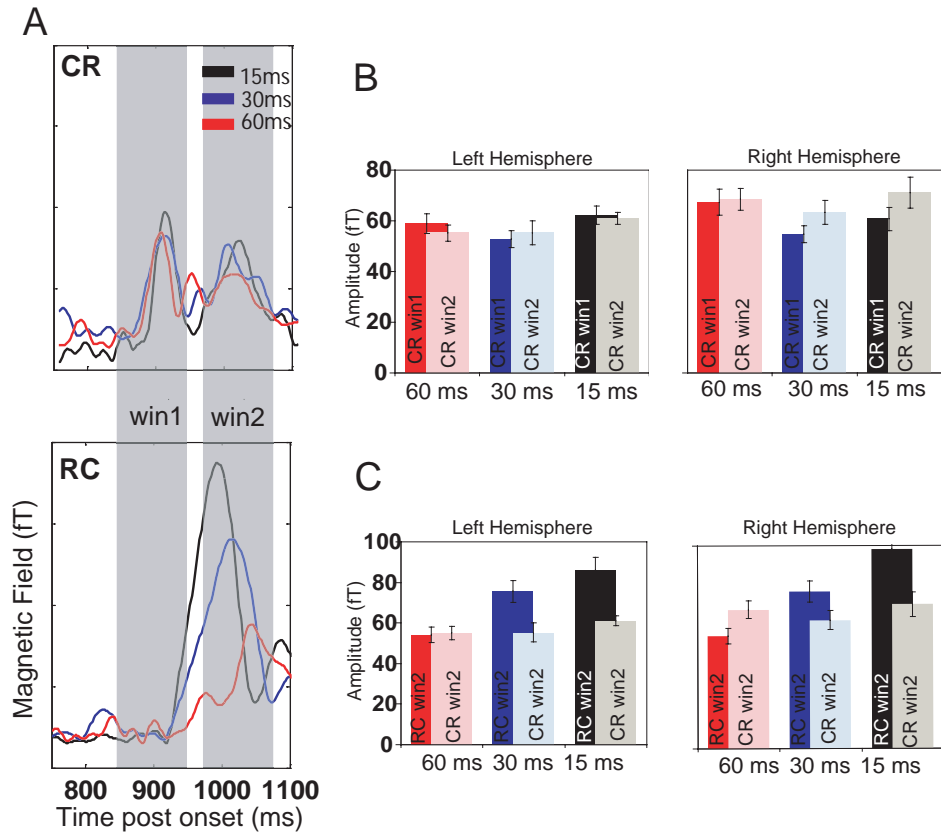
Figure 3.13: Transition response properties. A: responses to CR (top) and RC (bottom) stimuli for each of the pip-size conditions. B: comparison of CR peak amplitudes in win1 (dark colors) and win2 (light colors) in the left and right hemispheres. C: comparison of RC peak amplitudes (dark colors) and CR peak amplitudes (light colors) in win2.

the CR case, where no amplitude differences were found between pip-size conditions, the responses to the different RC shifts differ in amplitude such that transitions in the 15ms tone pip condition evoke the largest response, and transitions in the 60ms condition evoke the smallest amplitudes, the opposite trend to that seen in the onset responses to RC (Figure 3.11).

In addition to the amplitude differences, it is clear from Figure 3.13 that the responses to the different RC conditions occur at different times: the 15ms condition evoking the earliest response, and 60ms condition resulting in the latest. Figure 3.14 summarizes the latency data. The latencies of each of the random-to-constant conditions (Fig 3.14C light grey bars) differed significantly (Right Hemisphere: $F(2,46)=40.574$, $p<0.0001$; Left Hemisphere: $F(2,46)=38.342$, $p<0.0001$). Moreover, after correcting for pip-size (subtracting the appropriate pip size from the latency data for each subject; see Figure 3.10) the latencies were about 150ms and did not differ between conditions (Fig 3.14C dark grey bars). The difference between the mean uncorrected and corrected latencies was significant for each pip size condition (df=23; Left Hemisphere: 60ms t=10.34 $p<0.0001$, 30ms t=9.005 $p<0.0001$, 15ms t=2.61 $p=0.015$; Right Hemisphere: 60ms t=12.37 $p<0.0001$, 30ms t=4.54 $p<0.0001$, 15ms t=2.54 $p=0.018$). In the constant-to-random condition, the latency of both peaks (win1 and win2) did not differ between pip-size conditions (fig 3.14A and 3.14B). This is intuitively expected because the detection of change as the stimulus shifts from constant to random is based on the change in frequency and does not depend on pip size. On the other hand, in order to detect the transition in the RC stimuli, a listener has to wait a certain amount of time that depends on pip-size (an ideal
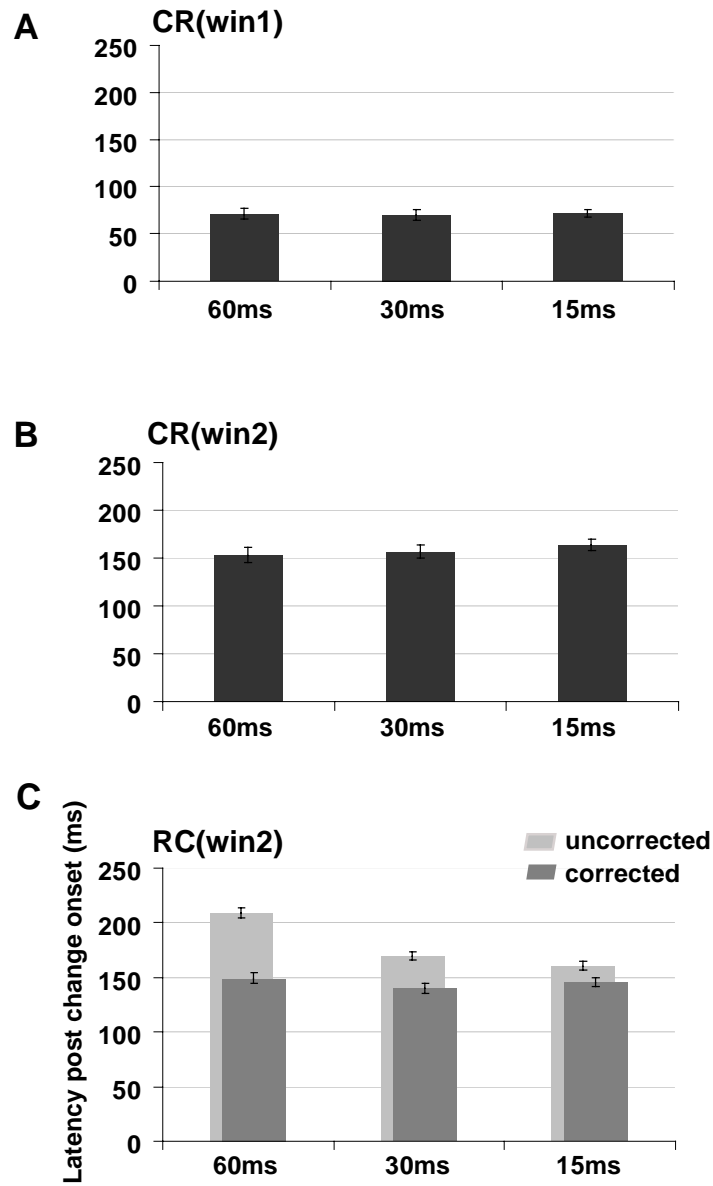
Figure 3.14: Transition response latencies. A: CR peak latencies in win1, B: CR peak latencies in win2, C: RC peak latencies.

observer has to wait one pip-size, see figure 3.10). Our data indicate that indeed the latency of the response to change was accurately adjusted according to stimulus statistics (pip-size), even though subjects were not actively trying to detect the change. A follow-up experiment with 6 additional participants used 60ms and 120ms pip durations, as well as a 'constant-to-constant' condition (see below). Stimulus generation and procedure were otherwise identical to the main study. We observed that responses to RC transitions with 120 ms tone pips exhibited a corrected latency of 150 ms, consistent with that observed for shorter pips. This indicates that auditory cortex is able to accurately tune to the pip-size for short (15ms) and longer (120 ms) conditions even when this is not a task-relevant parameter.

One aspect of the response to 'constant-to-random' stimuli that is dependent on pip-size is the emergence of a third peak ('b' in Figure 3.15, and also visible in Figure 3.13A), between the win1 and win2 peaks. This peak, with a dipolar distribution similar to that of the stimulus onset responses, (and of the response to change in RC stimuli) is not prominent for the 15ms condition but is visible in the 60ms condition. In figure 3.15, we replot the data for the 60ms and 15ms conditions in comparison with the response evoked by a 'constant-to-constant' (CC) transition (green line; acquired in the follow up study described above). Since the number of participants in the main and follow up studies is substantially different, it is hard to compare the amplitudes of the responses. But what is clear from figure 3.15 is that in terms of response dynamics, as pip-size increases responses to constant-to-random stimuli resemble responses to constant-to-constant transitions. The difference between the 60ms and 15ms conditions parallels the differences seen at onset (figure

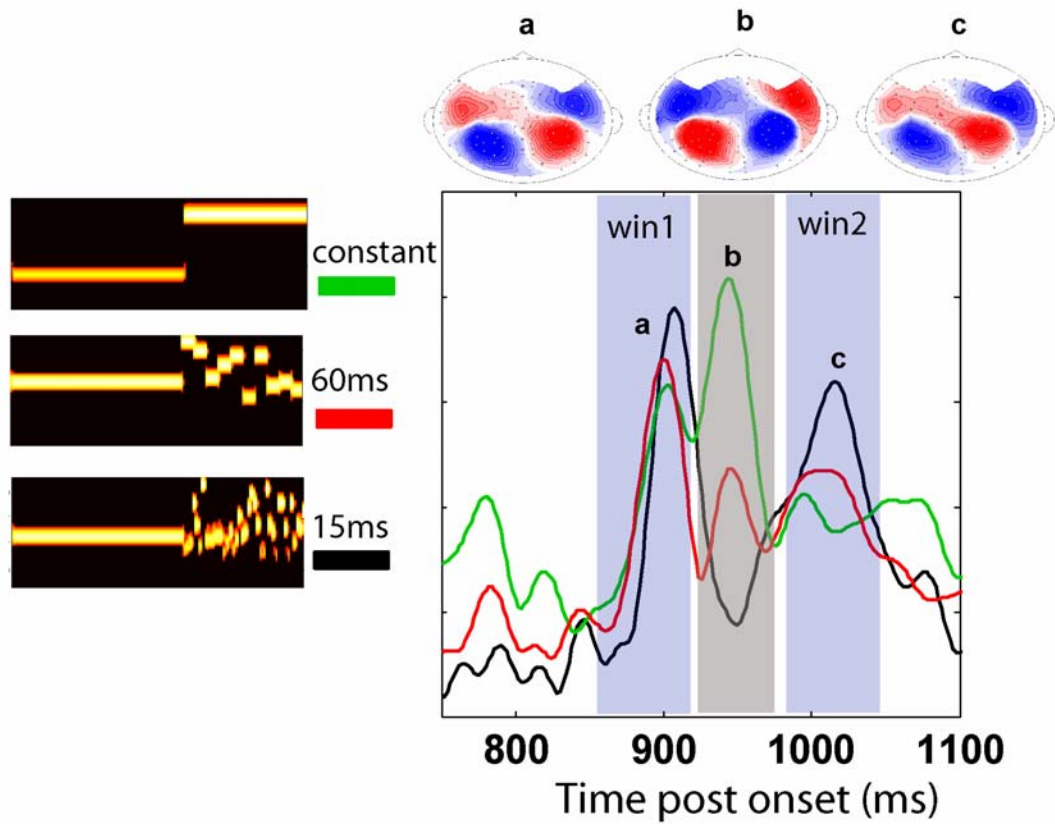**Figure 3.15:** Emergence of a third peak ('b'), with an M100 dipolar distribution, between the peaks in win1 and win2. The M100 peak is not visible in the 15ms condition (black) but emerges in the 60ms condition (red). Thus as pip size increases, responses to CR stimuli, resemble CC stimuli (green). The dipolar distributions corresponding to each of the three peaks ('a','b' and 'c') are also plotted.

3.12) and may reflect the same mechanisms. What is noteworthy, however, is that our data suggest that transitions from constant to constant stimuli (figure 3.15) are more similar to transitions from constant to random, than to transition from random to constant. This is discussed further below.

Behavioral experiment:

Except for one individual who failed to perform the task and was therefore excluded from analysis, the average number of misses, out of 240 targets was about 8 (3.3%) and the average number of false positives was about 4.5 (1.87%).

A repeated measures ANOVA on the false positive scores revealed a main effect of pip-size ($F(2,42)=9.1$ $p=0.001$), such that listeners had significantly less false positives in the 15ms than the 30ms or 60ms pip-size blocks (30ms: $t=3.23$ $p=0.004$; 60ms: $t=4.613$ $p<0.0001$), and a main effect of pattern ($F(1,21)=26.4$ $p<0.0001$). Subjects made significantly more false positive answers to the 'R' than 'C' stimuli (as might be expected).

A repeated measure ANOVA of the miss scores showed only a main effect of pattern ($F(1,21)=50.177$ $p=0.0001$): CR transitions were consistently missed more often the RC transitions (df=22; 60ms block: $t=4.3$ $p<0.0001$; 30 ms block: $t=3.6$ $p<0.000$; 15ms block: $t=6$ $p<0.0001$). This outcome is somewhat surprising since, as discussed above, at least for an ideal observer, detecting the transitions in the constant-to-random stimuli should be easier than the opposite transition. This is discussed further below.

The purpose of the behavioral experiment, conducted with the same stimuli and subjects as the MEG study, was to investigate to what extent brain responses are related to listeners' perception of the stimuli. Because the behavioral data and MEG data were not acquired simultaneously, there is a limit to the parallels that we can draw. We note, however, that average response times in the MEG decoy task correlated with average response times in the behavioral study (Pearson correlation coefficient r=0.64 *p*=0.001) indicating that subjects were at a similar functional state.

Subject response times, for all conditions, are shown in Figure 3.16A. A repeated measures ANOVA revealed only a main effect of pip-size (F(2,42)=29.283, *p<0.0001*) which was due to responses (in both CR and RC stimuli) in the 60ms block being generally slower than those in the 30ms (*p*=0.018) or 15ms (*p*<0.0001) blocks. This effect persisted even after correcting the RC stimuli by one pip-size (F(2,42)=11.13 *p<0.0001)*. The behavioral data thus differ from the properties of the change detection responses in 3 fundamental aspects: (1) MEG latency to a CR transition does not depend on pip size (Fig 3.14A), but behavioral latency is shorter for 15ms than 60ms pip size ;(2) Corrected MEG latency to a RC transition does not depend on pip size (Fig 3.14C), but behavioral latency is shorter for 15 than 60ms pip size; (3) MEG latency to a CR transition is about 80ms shorter than to an RC transition, but behavioral latencies do not differ significantly between these conditions.

A possible explanation for this pattern is that listeners define their 'behavioral integration window' based on the stimulus that requires the longer response , thus
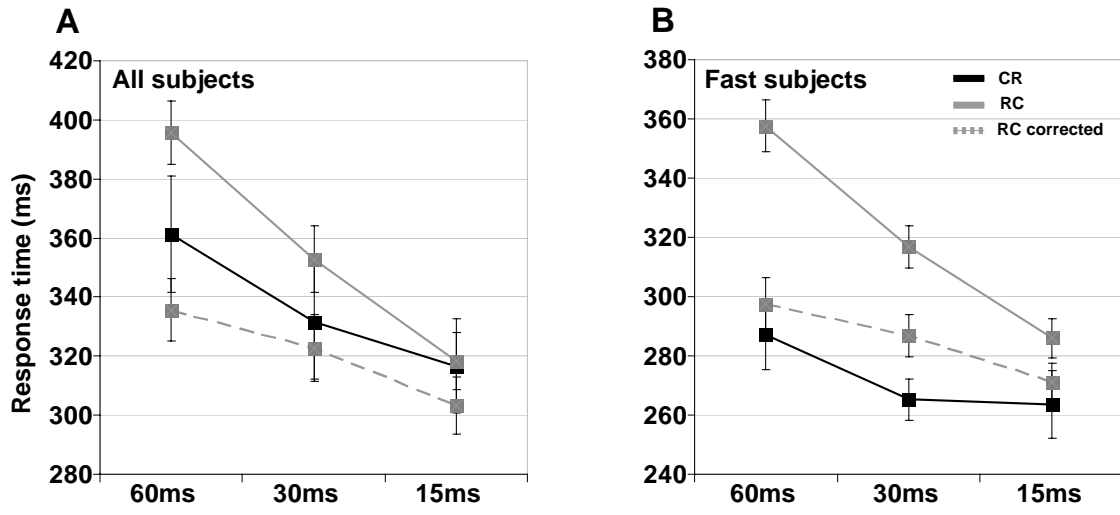
Figure 3.16: Behavioral response times. A: for all subjects B: for a subset (N=12) of fast subjects.

behavioral response times are 'bottlenecked' by the response time for RC stimuli. In order to investigate this possibility we analyzed the responses of fast and slow subjects separately. It has previously been noted (e.g. Dupoux, 1993) that response times of fast responders more closely reflect constraints of early processing, whereas those of slow responders may be affected by other factors.

Fast vs. slow responders

We divided the subjects into two groups: 'fast' (N=12) and 'slow' (N=11) based on their average response times. Mean responses times were 296 ms in the 'fast' group and 402 ms in the 'slow' group. A re-analysis of the behavioral data shows that fast subjects had significantly more false positives than slow subjects in the R condition ($F(1,21)=6.75$ $p=0.017$), and overall significantly less misses ($F(1,21)=10.2$ $p=0.004$). However, as in the All-subjects analysis, fast subjects consistently missed more constant-to-random transitions than random-to-constant ones ($F(1,11)=15$ $p=0.003$).

Figures 3.16B shows the response times of the fast subject group, in all conditions. A repeated measures ANOVA revealed main effects of pattern (constant-to-random transitions being responded to faster than random-to-constant transitions; $F(1,11)=89.2$ $p<0.0001$) pip-size ($F(2,22)=14.9$, *p<0.0001*) as well as an interaction of pattern×pip-size ($F(2,22)=11.9$ $p<0.0001$). The interaction is due to the fact that CR responses in the different pip conditions did not differ from each other, whereas RC responses did differ. After correction (subtracting one pip size from the response times to the random-to-constant stimuli), the main effects were smaller

(type: $F(1,11)=6.6$ $p=0.026$; condition: $F(2,22)=4.1$ $p=0.03$) and the interaction was gone.

Even though fast subjects exhibited somewhat different behavior than slow subjects, their pattern of responses still differed from what might be expected based on the brain data. Specifically, whereas first MEG response latencies differed by about 80 ms between conditions, behavioral response times differed only slightly. This suggests that later processes affected subject's conscious detection of change. Additionally, there was no interaction between subject group and any of the brain response measures (i.e. fast responders' electrophysiological responses to change did not differ from those of slower responders). Further evidence for the indirect relationship between the 'change detection' brain responses is that the subject who was excluded from behavioral analysis because of inability to accurately detect the changes in the stimuli (despite genuine effort), exhibited a pattern of MEG responses consistent with those of all other subjects.

It is noteworthy that these behavioral results contrast with those found by Chait et al (2005) for changes in interaural correlation. In that study, behavioral performance paralleled MEG activation: responses to transitions from uncorrelated to correlated noise occurred about 80 ms later than the opposite transitions. Additionally, greater miss rates were observed for the uncorrelated-correlated transition (which parallels the random-to-constant transition here) unlike the opposite trend (more misses when detecting transitions from constant to random) which was seen in the current study. It may be that listeners based their detection strategies on different perceptual cues to change. For example, one possible hypothesis is that in

196

this experiment subjects were more sensitive to onset of a constant tone (in the middle of a stimulus) than its offset. If behavioral detection was somehow based on the mechanisms that generated the M100 response, which was quite faint in CR stimuli (Figure 3.15) but vigorous in the RC stimuli (Figure 3.13) then we might be able to explain the findings that subjects had consistently more misses in the former than latter (contrary to what is intuitively predicted based on statistics; see also DeWeese & Zador, 1998). This aspect should be investigated more methodically in future studies.

3.2.3 Discussion

Transitions between 'random' and 'constant' signals are processed in fundamentally different ways depending on the direction of the transition. The constant-to-random change evoked a sequence of cortical activations that differed from the opposite transition (random-to-constant), in time, and in the dipolar distributions (indicating different underlying neural machinery). Interestingly, the same asymmetry, with the same dynamics and topographies, was found in MEG responses evoked by transitions between interaurally correlated and uncorrelated noise signals (Chait et al, 2005). That study interpreted the asymmetry as reflecting different requirements for binaural processing (specifically, more temporal integration is required to detect a transition from interaurally uncorrelated to correlated). However, the similarity between response patterns in that study and this one forces us to widen the scope of the interpretation. The asymmetry might reflect more general processing constraints.

The stimuli of the two studies differ acoustically (narrowband vs. broadband, monaural vs. binaural, stationary vs. fluctuating), and the perceptual attributes of the change are different, but they share the same abstract characteristic of a transition between disorder (or fluctuation) and order (or constancy). The similarity of brain responses observed in both studies suggests that these responses reflect a general mechanism responsible for detecting transitions between these states.

As discussed in the introduction, an observer presented with an RC stimulus needs time after the onset of the constant tone in order to discriminate it from the onset of another pip. For equal-duration pips, this time is *at least* one pip duration. It is remarkable that MEG response latencies to an RC transition indeed follow this trend: when corrected by subtracting one pip duration, they are constant (150 ms from change onset) and do not depend on pip size. Thus, the preattentive mechanism that this response reflects adapts accurately to durations from 15 to 120 ms long, even though the subjects were not actively listening to the change in the stimuli. We chose pip sizes in the 15-60 ms range to bracket a value (30 ms) that has been proposed as the size of the putative cortical short-term integration window (Poeppel, 2003; Wang et al, 2003). On this basis one might expect responses to 15 ms and 30 ms stimuli to exhibit the same response pattern. Such was not the case: latencies for a pip size of 15 ms were significantly shorter than for 30 ms. In light of these findings it is interesting to use yet smaller and larger pip sizes in order to investigate the limits of this cortical adjustment. A possible caveat is that in the current experiment different pip sizes were blocked separately and different response patterns may be observed if

the presentation is completely randomized. This will be investigated in a future experiment.

Change detection in humans is usually investigated with the MMN paradigm (Polich, 2003), which is based on comparing brain responses to deviant (low probability) signals presented among standard (high probability) signals. This technique has been instrumental in revealing the properties of sound that auditory cortex is sensitive to. The MMN response (derived by subtraction between the responses to standard and deviants) peaks between 100-200ms. It is interpreted as reflecting a process that registers a change in a sound feature and updates the ongoing representation (Winkler, 1996).

In mismatch experiments stimuli are usually short and separated by silent intervals but, as discussed in the introduction, in natural environments listeners are subjected to continuous auditory stimulation. Changes are superimposed on the continuous waveform that enters the ear and a listener thus needs a mechanism to decide at which point in a *continuous* sound the change is introduced. It is such a mechanism, we believe, that we are tapping into with our paradigm. The responses seen here commence at about 50 ms post change, which is much earlier than the latencies registered with MMN, but include also later responses at about 100-150 ms that might be related to those observed in MMN.

As discussed in the introduction (see also legend of Figure 3.10), an ideal observer presented with these stimuli would be able to respond instantaneously to a CR change, because the appearance of the first tone pip in the random segment constitutes a deviation from the acquired representation of the initial segment. In the

case of the reverse transition, from random to constant, an ideal listener would have to wait slightly more than one pip size in order to differentiate the stimulus from the 'Random' condition. That is, if corrected by pip size, latencies to constant-to-random and random-to-constant stimuli should not differ. Our data indicate that the auditory cortices of human listeners do *not* behave like ideal observers. Even after correcting for pip size, the shift from constant to random was detected 80ms faster than the transition from random to constant. Additionally, the pattern of results imply that, even after correction for pip size, constant tones are detected faster if preceded by a constant tone, than if preceded by a random sequence of tones (see Figure 3.15). It is clear therefore that the cortical mechanisms which underlie the MEG responses we record, are not detecting change, per se. Our observations may however serve to construct hypotheses about the computations that these mechanisms reflect and how they are related to perception.

The brain responses to transitions that we observed share characteristics of well-known components of auditory cortical responses to sound onsets. MEG/EEG onsets commonly evoke a sequence of three peaks at around 50 ms (M50, analogous to P1 in EEG), 100 ms (M100, analogous to N1 in EEG) and 150 ms (M150, analogous to P2 in EEG) post onset (Naatanen & Picton, 1987; Hari, 1990; Roberts et al, 2000; Yvert et al, 2001). Depending on the study, the M50 and/or M150 peaks may be absent (below we provide a possible explanation), though the M100 (or N1 in EEG) is typically seen as a vigorous response (but see Chait et al, 2004; 2006), and is thought to be related to a system that monitors abrupt changes in sensory input (Naatanen & Picton, 1987).

The onset responses to our stimuli (Figure 3.12) were consistent with this pattern (strong M100, weak M50). Responses to transitions, however, depended on the order of the transition (RC vs CR). The CR transition (and also CC) evoked a vigorous response at about 60 ms post-transition with a topography similar to M50. RC stimuli evoked a robust response at about 150 ms after the logical transition (physical transition + pip size), with a topography similar to M100. Brain responses are thus highly asymmetrical according to whether the order of the change is from random to constant, or constant to random. Similar response patterns have been reported for CC signals such as transitions between synthesized vowels (Martin & Boothroyd, 2000), for abrupt changes in the pitch or timbre of continuous complex tones (Jones & Perez, 2000) and for transitions from monophonic to pseudo-stereo sound (Ross et al, 2004).

Based on these observations, and the fact that we found a very similar asymmetry in responses to changes between interaurally correlated and uncorrelated noise stimuli (Chait et al. 2005), we can speculate about the functional roles of the mechanisms generating the peaks. The first deflection at ~70 ms post change that occurs in situation when the change is immediately detectable (such as at the shift from constant to random or from constant to constant) may be reflecting the operation of the 'obligatory cortical integration window'. The response at around 100-150 ms with the M100 like dipolar distribution may reflect the operation of another, 'adjustable window', provided by a separate neural substrate, which integrates incoming information to reach a sufficient level of certainty that a change has occurred. At a more abstract level, these two windows can be thought of as reflecting

construction and destruction of perceptual representations: The M50-like deflection at ~70 ms post change that occurs in shifts from constant to random (or from constant to constant) may be reflecting the *detection of a violation of* a previous representation of an ordered object. The response at around 100-150 ms with the M100 like dipolar distribution may reflect the *construction* of a new representation, the exact latency of which would depend on how hard it is to construct the representation (see also Gutschalk et al., 2004b). Note that the fact that onset responses to constant tones after another constant tone (CC transition) are faster than to constant tone after a random sequence (RC transition) indicate not so much a difficulty in constructing the representation (as the final representation is the same in both cases) but differences in the mechanisms leading to the construction and how fast they kick in. The inherent latency difference between how fast representations are constructed and how quickly they are destructed is reminiscent of the words of B. Julesz (Julesz & Tyler, 1976): *"...many perceptual phenomena require more effort to go from a disordered state to an ordered one than vice versa"*. Onset responses (the classic M50, M100, M150) might be interpreted in similar terms, as resulting from the construction of a representation of a new 'order' (with additional activation related to energetic onset).

Our stimuli, whether constant/random tones or correlated/uncorrelated noise, are very simple examples of 'order' and 'disorder'. More experiments are needed to map out the properties of these responses, and resolve speculations as to their role. For example, replicating these experiments with regularly alternating pip sequences (rather than random) would determine the respective importance of 'randomness' vs 'variation' in the phenomena we observed. We speculate that more complex stimuli

that are characterized by regularity would fall under the 'order' category and that other stimuli involving transitions between regularity and irregularity should evoked similar response patterns to the ones observed here. There is in fact evidence, from the pitch literature, that this is indeed the case. In particular, the response to the onset of periodicity in click trains (Gutschalk et al., 2004a) or noise-like stimuli (Krumbholz et al., 2003) shares characteristics of the RC transition response observed here, or the uncorrelated-to-correlated response of Chait et al. (2005).

A recent discovery of am MEG pitch onset response (POR) is generating considerable interest (Gutschalk et al., 2002; Krumbholz et al, 2003; Gutschalk et al., 2004a; Seither-Preisler et al., 2004; Ritter et al, 2005; Chait et al., 2006). These responses, hypothesized to reflect cortical pitch processing mechanisms, are evoked by transitions between irregular click trains, which do not have a pitch and regular click trains which are perceived to have a sustained temporal pitch (Gutschalk et al., 2004a) or by transitions between white noise and iterated rippled noise (IRN; Krumbholz et al., 2003). The evoked responses are interpreted as related to pitch because all other aspects of the stimuli (spectral content, sound energy) are not altered in the transition. However, another possible way to describe the transitions is as transitions from randomness to regularity. The POR responses indeed exhibit properties similar to the ones observed here and in Chait et al (2005) for random-to-constant stimuli: These responses are characterized by a sole peak at about 150ms post change with an M100 dipolar distribution with a latency and amplitude that depend on stimulus statistics (Krumbholz et al, 2003; Seither-Preisler et al, 2004; Gutschalk et al, 2004a; Chait et al, 2006). Interestingly, when investigating PORs

evoked by transition from one IRN to another (Ritter et al, 2005) the pattern evoked by the transition changes, the response develops an M50 component that precedes the POR (see Figure 8 in Ritter et al, 2005). This is entirely consistent with a 'constant-to-constant pattern' observed here (Figure 3.15). It is tempting to conclude that this response is not specific to pitch processing mechanisms per se, but to a mechanism that handles transitions between states that differ along a more abstract dimension, such as degree of regularity or order (see also Hall et al., 2005).

If our hypothesis is confirmed in future experiments, the investigation of change responses can contribute to auditory science in a number of ways: it can provide a tool to measure what is deemed 'regularity' by the auditory system (analogous to the role of the MMN in determining salient features of sound). It can also give us a key as to what aspects of ongoing stimulus statistics the system is sensitive to, and how it estimates them. The dependency of latency on stimulus statistics may allow us to probe the characteristics of temporal integration of incoming acoustic information. The preattentive nature of the responses (and, in Chait 2005, 2006 their dependency on binaural processing) may allow us to situate these stages within the processing chain. Future experiments should investigate changes along different dimensions (e.g. loudness, timbre, direction) as well as different statistical properties. Recent neurophysiological studies (Fairhall et al., 2001; Ulanovsky et al, 2003; Dean et al, 2005; see also DeWeese & Zador, 1998) are beginning to investigate how neurons adapt to changes in stimulus statistics. The paradigm introduced here may serve as the human non-invasive counterpart of this

research, and illuminate the dimensions of auditory signals that are relevant for the construction of perceptual representations.

3.3 Further issues

The experiments described in the previous sections revealed what appears to be a general mechanism that is sensitive to sudden changes in the statistical properties of stimuli. Sensitivity to change is a fundamental aspect of auditory perception: the information about the appearance or disappearance of objects in the environment is reflected in changes in the statistical properties of the waveform that arrives at the listeners ears. We showed that order-to-disorder transitions appear to be processed by a different neural population than disorder-to-order transitions, and with different temporal dynamics. The latencies of the responses to the two conditions are indicators of the size of the windows over which information is integrated (or gathered) in each case.

3.3.1 A separate mechanism for transitions from 'order' to 'disorder' and vice versa?

The striking conclusion of this pair of experiments is that symmetric transitions between 'order' and 'disorder' in both interaural correlation and frequency domains transitions evoke activity within different cortical circuits with different temporal dynamics. The conclusion about distinct neural substrates is based on the different polarities of the magnetic field. However, there is additional evidence, from brain-injured patients that this is indeed the case. Jones (2002a) recorded EEG responses from the brain of a comatose patient in responses to pitch change of continuous tones

(comparable to CC stimuli here) and to the end of oscillatory pitch changes (comparable to RC stimuli here). He reports that the transition response to the CC stimulus was similar to that of controls, but the response to RC stimuli was absent.

3.3.2 Binaural sluggishness

'Binaural sluggishness' (Grantham & Wightman, 1979) refers to the apparent insensitivity of the binaural system to rapidly varying interaural configuration: In a variety of tasks subjects consistently show a relative difficulty in detecting dynamic changes in binaural cues (Grantham, 1995). For example, when an ITD of a wide band stimulus alternates between -500µ and +500µ, listeners report the position of the image as moving between one ear to the other and back. However, for rates as low as 2.4 Hz subjects cannot follow the full excursion of the image and when the rate increases to about 10-20 Hz no movement at all is perceived (Blauert, 1972; Grantham & Wightman, 1978). Other studies have also demonstrated reduced (relative to monaural dynamic variations) binaural sensitivity to step changes in interaural phase. The threshold of a short $S\pi$ probe tone in the presence of a N0 masking noise that was preceded by a $N\pi$ masking noise is increased (detection worsens) as a the tone onset is moved closer to the $N\pi$ noise offset (Kollmeier & Gilkey, 1990). These data indicate that listeners compute the effective interaural phase value over a binaural integration window that in turn influences detection in binaural unmasking situations. This integration appears to be substantially bigger (about 3 times larger) than in comparable monaural situations (where the interaural phase of the noise masker is kept constant but the level is reduced; Kollmeier &

Gilkey, 1990). Similar findings have been reported for detection of dynamic changes in interaural correlation (Culling & Summerfield, 1998; Boehnke et al, 2002).

'Binaural sluggishness' is assumed to reflect the existence of a 'binaural integration window' that operates subsequent to the site of binaural interaction, i.e. at or above the IC. However, Joris et al. found no evidence of sluggishness in responses of IC cells to stimuli with dynamic interaural correlation (Joris et al, 2006). As suggested in chapter 3.1 here, it is possible that the mechanisms we observe reflect the operation of this integration process. However, the similarity between the tone (chapter 3.2) and IAC (chapter 3.1) data point to the fact that this mechanism is not special to binaural processing. Our data and the ensuing discussion suggest that 'Binaural sluggishness' might not be a result of special 'binaural integration windows' but may in fact result from the same integration mechanisms that process monaural data. The apparently larger integration in the binaural case may stem from the way in which binaural information is represented. By searching for a monaural stimulus (for example by adjusting the pip size/distribution properties of our tone stimuli) that results in the same response properties (peak latencies) as the binaural stimuli in the IAC experiment, we may be able to learn more about how binaural information is centrally represented.

3.3.3 Integration windows

In Experiment 2 (where the dimension of change was frequency), pip duration was varied as a parameter. This allowed us to investigate how the size of the integration-window used for change detection (How long does it take for change to be detected?)
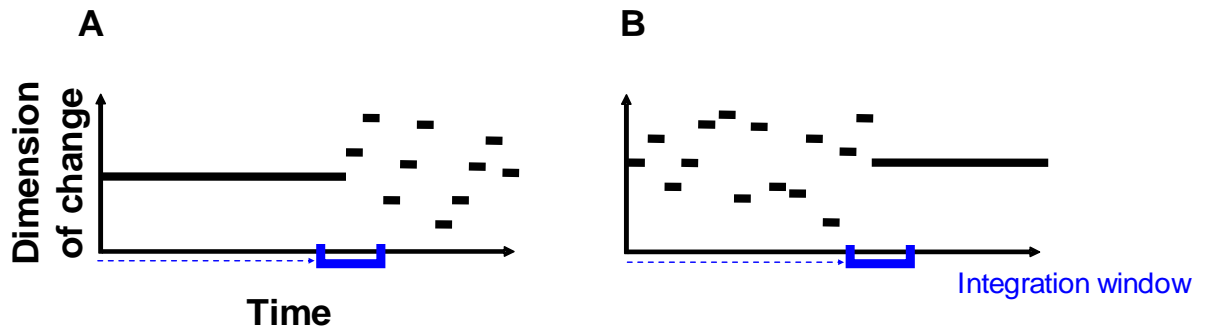
Figure 3.17: Schema of the stimuli used in our change detection studies. In order to detect the changes, the auditory system uses an integration window that extracts statistical properties from the ongoing signal. In B a longer integration time is required to infer that there has been a change. With MEG, we are able to measure the latency of brain responses to change and investigate the properties of these temporal integration mechanisms: how adjustable are they? How fast does adjustment occur? Does adjustment require attention or does it happen automatically?

is adjusted according signal statistics (longer integration windows are needed for longer tone pip duration). By measuring brain response latency as a function of this parameter, we can estimate the amount of temporal integration that precedes the site that generates these responses (see Figure 3.17). We showed that this adjustment occurs very accurately for all pip durations used in the study (15ms-120ms) even when subjects are not actively attending to the stimuli. However, in the current different pip durations were blocked separately and this might have affected the response acuity. Now that we have established that transition related responses are measurable and are adjustable to the statistics of the stimuli, it is important to examine the effect of a randomized presentation on the precision of cortical adjustment.

3.3.4 Offset responses

Offset responses are commonly described in the literature as morphologically similar to onset repsonses (i.e consisting of M50, M100 and M150 deflections). However, the offset responses in the tone experiment exhibited properties similar to the transition-related responses: Whereas R (and CR) sequences elicited no discernible offset responses, whatever the pip size, C (and RC) sequences elicited a vigorous response at 150ms post onset with an M50 like distribution (similar to the peaks in the RC transition). The similarity of responses indicates that, as far a MEG responses go, the transition from 'constant' to silence may be similar to the transition from 'constant' to 'random', which we interpreted above as "a detection of a violation of a previously acquired representation". However this is in contrast to the offset responses in the

IAC experiment stimuli which, for both correlated and uncorrelated noise, exhibited an offset pattern that is more similar to what is described in the literature (a M100-like deflection at about 100ms post offset). This indicates that change-related responses cannot be explained solely in terms of offset responses.

A cursory review over all my experiments (for which offset responses were not examined in detail) suggests that offset responses, like onset responses, may be dependent on context and task. In contrast to the abundance of literature about onset-responses, I am aware of only a handful of systematic investigations of the off-responses (Noda et al., 1998; Pantev et al., 1996; Hillyard & Picton, 1978). As these dependencies may produce large effects, and the conditions under which they appear are not well known, it is important in future studies to explore the dependencies more fully. They may in fact provide a new key to understand underlying processes.

3.3.5 At what processing stage do change detection mechanisms operate?

A previous study (Chait et al, 2005; Chapters 2.1 and 2.2 here) employed stimuli with simultaneous changes in two dimensions (change in interaural correlation, appearance of a tone; see Figure 2.12 here). We observed two *separate* response peaks, one for each dimension. This might indicate that the change detection mechanisms operate on a representation in which the noise and tone are already segregated. Howerver, this observation appeared as a side-effect in the experiment, and it needs to be investigated more rigorously with appropriately controlled stimuli. The question is vital to the study of scene analysis. Since the change detection mechanisms appear to

be automatic (not require attention), finding that they operate on segregated streams would imply that this segregation too is automatic.

3.3.6 Do these mechanisms require attention?

In these experiments, brain responses were recorded while subjects performed a task irrelevant to change detection. The fact that the responses occurred even when they are irrelevant to the task, does not mean that they are unaffected by the task. In order to investigate whether these mechanisms are indeed completely automatic or depend on general attentional/computational resources, it is crucial to study how a subject's attentional load (an easy, or a very hard irrelevant task) affects these mechanisms.

Kubovy (1981) used a stimulus which consisted of concurrent pure tones. He reports that when one of the tones was lowered in intensity and then restored to its original loudness value (similar to CC stimuli here) its pitch dominated the perceptual experience of the listener as though that tone became the figure and the rest of the concurrent tones were relegated to background. One functional role for the change-related responses we observe may be attentional capture – diverting attentional resources to the newly detected event, which may result in the perceptual effect described by Kubovy. This may be investigated in future studies by examining the effect the non-attended change has on subject's performance in some other task.

3.3.7 Transition-related MEG responses and behavior

Although the electrophysiological signature of the transitions between random and constant tones and between interaurally correlated and uncorrelated noise was similar,

212

the behavioral response patterns differed significantly. The difference in the behavioral patterns was manifested both in response time difference between conditions and in miss-rate patterns.

In the IAC experiment, subjects behavior correlated with brain responses (and with what would be predicted by an observer model) in that they responded significantly faster to 1→0 transitions than to 0→1 transitions. The difference in response time – 80ms—was similar to the electrophysiological latency difference between the first responses to the 1→0 and 0→1 conditions. In the tone experiment subjects exhibited essentially no difference in response time between CR and RC conditions, even though brain responses to CR conditions preceded those of RC conditions by about 80ms.

In a analogous manner, miss rates in the IAC experiment were consistent with predictions based on a theoretical observer in that 0→1 transitions were missed more often than 1→0 ones. However, in the tone experiment we observe the opposite pattern: CR transitions (which are theoretically easier to detect) are missed significantly more often then the, opposite, RC transitions. This is true both for fast and slow subjects.

There is no straightforward explanation for the difference in behavior. A possible speculation may be that the disparity in performance is a result of subjects using different perceptual attributes of the stimuli to detect the change. Participants in the IAC experiment seem to have based their judgment on a percept of narrowing (in the case of 0→1 stimuli) or widening (in the case of 1→0) of the source: in post experimental interviews, when asked to describe the change in the stimuli subjects

tended to describe it as *movement* of the source. 0→ transitions were reported as movement towards the center of the head, whereas 1→ transitions were described as a single focused source that is "stretching" and moving away from the center towards the two ears. The fact that the interaural correlation change was perceived as gradual, even though the physical change was abrupt, may be an indication for the existence of a sliding binaural temporal integration window (Culling & Summerfield, 1998; Akeroyd & Summerfield, 1999; Boehnke et al., 2002) over which the perceived IAC value is computed. That indicates that the subjects might have based their performance on a percept that is directly linked to the outputs of the integration windows whose activity might be reflected in the transition-related MEG responses.

On the other hand, a possible explanation of the behavioral pattern in the tone experiment is that subjects were more sensitive to onset of a constant tone (in the middle of a stimulus) than its offset. The miss-rate pattern as well as the incongruence between MEG and behavioral responses may be explained if behavioral detection was somehow based on the mechanisms that generated the M100 response, which was quite faint in CR stimuli (Figure 3.15) but vigorous in the RC stimuli (Figure 3.13). In the concluding section of this dissertation I will attempt to justify this proposition.

In section 2 of the previous chapter, we already discuss the fact that these early cortical mechanism are probably reflecting the extraction of the target and that this process is independent from the mechanisms that bring the change to the awareness of the listener. The investigation of stimuli such as the ones here may shed more light on the relationship between perception and early brain activation.

3.3.8 Further experiments

The investigation of these general change detecting mechanisms is fascinating because it may shed light on the statistical heuristics with which our brains sample, represent, and detect changes in the world, including changes that are not the immediate focus of attention.

Future experiments are needed to test the generality of the present findings, to clarify the relation between behavioral and brain responses to change, to better understand the rules that determine factors such as integration time, and to characterize the type of changes that are highly detectable by the brain, compared with those that are less detectable. It is interesting to use other change dimensions (for example loudness, pitch, location) and to investigate more diverse distributions (so far we used only constant↔random changes) in order to address issues such as: what kinds of changes are observable by the brain and what kinds are not? What does that imply about the computations that are taking place and the way in which sensory information is represented? Are there stimulus changes that are not detected behaviorally yet still evoke change-related brain responses?

In addition to studying the dynamics of change detection, the paradigm introduced in these studies may be used as a *methodological tool* to investigate which dimentions of auditory signals are perceptually relevant: For example, are changes in different features, which are relevant to auditory objects (loudness, ITD, ILD, pitch) processed in the same way and equally quickly? This may shed light on their relative importance in determining the emergence of auditory objects.

Chapter 4: Edge detection

4.1 Change detection

In natural environments, comprised of many concurrent sound-generating objects, survival depends on listeners' ability to segregate the single complex waveform arriving at each of the ears into separate source components. Surprisingly, Eramudugolla et al (2005) report that listeners are remarkably poor at detecting the disappearance of an individual auditory object when listening to scenes containing more than four objects, unless their attention is pointed to the identity of a potential change. The stimuli in those experiments consisted of two versions of an auditory scene (a virtual space comprising several auditory objects) from the second of which one object might be missing. The two versions of the scene were separated by a 500 ms segment of white noise, which the authors justify as necessary to mask any transients or echoic memory that might cue the listener's attention to the change. However, what the results indicate is that continuity of the scene (which was disturbed in the current stimuli by the noise bursts) is an extremely important cue that most likely subserves listener's performance in natural conditions. The stimuli used in the series of experiments reported in this dissertation were designed to tap these natural change detection processes.

Chapter 2 examined the processes by which tone onsets are detected amidst background noise. We identified a cortical deflection, occurring at about 150ms post tone onset which corresponds to the segregation of the tonal target from the

background. These mechanisms were similar for the two kinds of targets we tested –
HP and TN-- perceptually alike but physically very different signals, indicating that at
about 150ms post onset the targets have already been mapped to similar
representations in cortex. We found further that, when a change is introduced
simultaneously to two aspects of the stimulus (change in the noise, appearance of a
tone), two distinct responses are elicited, each similar to the response to one of the
changes. This suggests that certain aspects of the segregation of the incoming signal
into its constituent components have occurred by 150ms post onset. The experiments
also strongly show that this early organization of the auditory signal is pre-attentive
and that later processes contribute to the way a listener appreciates the changes in the
environment and how fast s/he is able to detect them.

Nevertheless, at the end of Chapter 2 we argued that these responses cannot be
attributed to a fully general figure-ground segregation mechanism (for example such
as implied in Scheich et al, 1998), because their properties also depended on the
nature of the background (continuous tone, or noise). A more conservative hypothesis
is that the responses we observe reflect *change detection.* The emergence of an object
within a background is often signaled by the existence of "edges", or transitions in the
properties of the stimulus as one moves (in space for a visual scene, in time for an
auditory scene) across the sensory map. Perception of an object obviously requires
more than this, but one can hypothesize that edge detection is one of the fundamental
processes on which object detection – a more complex process – is built upon. In
Chapter 3 we show that indeed the cortical activation we observed in response to an

appearance of a tonal target in background noise is consistent with the interpretation that the underlying mechanisms reflect a process of change detection.

The "edge" that distinguishes an auditory object from the background may be defined by a change in power over time or frequency, or a transition of some other parameter such as frequency (for a narrow-band stimulus), spectral shape (for a wide-band stimulus), modulation frequency, etc. (Bregman, 1990). However, many natural soundscapes are already full of fluctuations, and indeed the particular fluctuations caused by the object (over space, time, frequency, etc.) may be the sole characteristic of the object that distinguishes it from the background. Alternatively, the background may fluctuate and the object be characterized by an absence of fluctuation. The relevant change that signals the presence of an object may then be a change in *statistics* (e.g. variance) of a characteristic such as frequency or amplitude. The appearance or disappearance of an auditory object in the environment is reflected in a change in the statistical properties of the ongoing sound that reaches a listener's ears (e.g. Bregman et al., 1994)

A fundamental property of perception is object invariance or *abstraction*: We can recognize a melody whether the pitch is high or low or identify a familiar object despite differences in view point. At the same time, when considering lower level processes related to analyzing the sensory scene it is clear that living systems are sensitive to context and to *relative* information: Observers are poor at determining the absolute brightness of light, or absolute pitch of a note, but are very accurate at detecting brightness or pitch difference. In a recent demonstration Demany & Ramos (2005) showed that although listeners are unable to tell whether a pure tone presented

after a chord of pure tones was a component of the chord, if the single tone following the chord was slightly higher or lower in frequency than one of the chord's components then listeners are able to perceive a pitch shift in the corresponding direction: that is, listeners more easily hear out a change in the components than the components themselves. This effect may very well be the explanation for the 'change deafness' observed by Eramudugolla et al (2005), and in fact it might be the case that the central nervous system is mainly based on forms of change detection (e.g. Barlow, 1961). Adrian (1946; Pp 83-85) addressed this issue:

> "The difficulty that we should find in recognizing a particular intensity of sound depends on a fundamental property of living systems, their power of adaptation to changed conditions. All living cells tend to come into equilibrium with their surroundings if they can, but the process takes time, and therefore a sudden change will cause far more disturbance than one which is established gradually….The adaptation of the sense organs is therefore an example of a general property of cells and organisms. …The sense organs, then, and the nerve-cells to which they lead, owing to their power of adaptation, are able to perform a process like differentiation in which absolute intensities become far less important than rates of change…."

I believe the cortical activations observed in the experiments reported here reflect such early representational levels. The IAC and tone-change stimuli used in Chapter 3 were physically and perceptually different, yet they shared the same abstract characteristic of a transition between 'order' and 'disorder'. A comparison between the two studies shows a striking similarity in the temporal dynamics of the transition response, revealing what appears to be a general mechanism that is sensitive to sudden changes in the statistical properties of stimuli. We showed that order-to-disorder transitions appear to be processed by a distinct neural population than disorder-to-order transitions, and with different temporal dynamics. The exact

219

latencies of the responses to the two conditions are indicators of the size of the integration windows used in the two cases.

The stimuli used in both studies were obviously particular and rather simple in terms of the feature dimension along which the change occurred, and the statistics of changes, and it is important to verify our conclusions for a more diverse set of signals. However data that emerge from MEG studies of pitch processing, reviewed in Chapter 3, are consistent with the interpretation that responses reflect transitions between different levels of regularity.

The stimuli that we used in Chapter 3 are conceptually similar to those used in studies on visual texture discrimination. Figure 4.1 (from Julesz, 1962) illustrates the immediate impression of two distinct fields that a difference in texture statistics can elicit. While it is easy to understand how subjects can use slight differences in the first order statistics (e.g. mean brightness) to delimit a visual object, performing the same operation on the basis of higher order statistics, as subjects appear to do preattentively, is a computationally more demanding task. In Fig. 4.1 (figure 5 from Julesz, 1962) -- the two areas in the figure have the same first-order statistics but are easily discriminable on the basis of their different second-order statistics. However, it is noteworthy that the transition is locally ambiguous, and a clear separation between the two patterns emerges only after integration of information across a certain extent. The perceptually salient boundary is a result of integration of global statistical information. In this sense, the process of detecting the visual edge is similar to what happens at a transition from interaurally-correlated to uncorrelated noise, or from random sequence to constant tone. That transition cannot be detected as a mere

transition along a feature dimension. Detection requires acquiring a representation of the statistics before the transition, comparing it with a representation of the statistics after the transition, and deciding whether the two are compatible with the absence of change, or instead indicate a transition. Integration requirements depend on the nature (statistics) of the patterns on either side; these differ according to the direction of change, and this justifies the asymmetry of response.

This comparison demonstrates that the kinds of operations needed, both in the visual and auditory domains when detecting transitions from constant to random or constant to constant (and in vision: between a white background and black square) are conceptually different from those that are required when going from random to constant or when transitioning between two different kinds of regularity, such as transitions between different IRN pitches (and in vision: between two statistically different patterns). In the former case all one needs to detect is a *violation* of a previously acquired regularity. In the latter case, change detection requires the *formation* of a new regularity. Our MEG data demonstrate that these operations indeed take different times and are based on different neural mechanisms.

At this point one might raise the 'philosophical' question regarding what it is that is being detected: In RC transitions, for example, are we detecting the *offset* of disorder or *onset* of order? I would like to speculate that the specific dynamics of our responses indicate that, depending on the nature of the transition, we are detecting different things: in CR or CC transitions, the first thing we detect (the M50-like peak) is the *offset* of order (destruction of a previously acquired 'order') in RC transitions, it is the *onset* of order that is being detected.
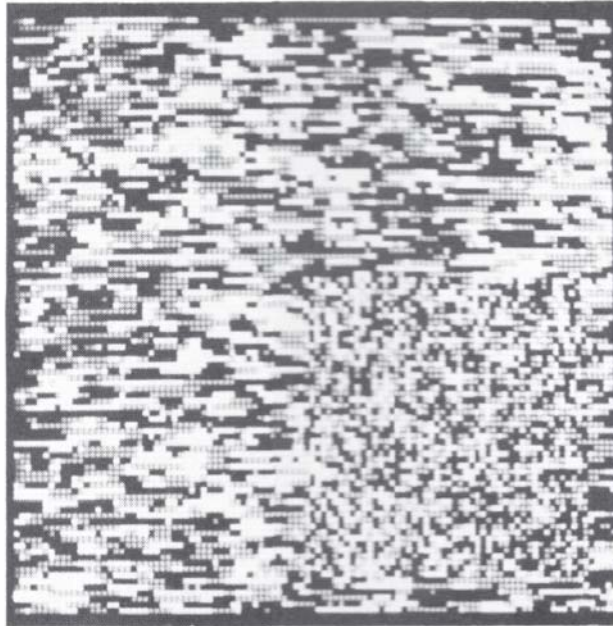
Figure 4.1: Figure 5 from Julesz (1962). Two random textures with identical first order probability (black, grey and white pixels of identical probability) but different second order probability: Any two pixels in the smaller field are statistically independent in their brightness value while the larger field is a Markov process. Observer's are able to automatically use this difference to segregate the two textures, though the boundary is fuzzy.

4.2 Auditory edges?

From the comparison with the visual domain, it becomes apparent that it might be appropriate to refer to the change-detection mechanisms as a form of *edge-detectors*. And our preceding discussion implies that there might be different kinds of auditory edges.

The extraction of edges from images is believed to be a principal objective of early visual processing (Hubel & Wiesel, 1965; Burr et al, 1989; Lamme et al., 1999) and a first stage in the computation of objects and scene analysis (Marr, 1982; Julesz 1971; Petersen, 1999). The feature-based theory of vision posits that the visual scene is first decomposed into edges, and that these are subsequently bound together to create visual forms. Similarity and proximity between extracted edges facilitate their perceptual integration (Marr, 1982; Treisman, 1985; Geisler et al., 2001). Visual edge detection is believed to be accomplished by sensitivity to local gradient in brightness or color, or else more complex textural changes such as in Figure 4.1 here (Marr & Hildreth, 1980).

Recent neurophysiological studies in the auditory domain support a similar role for primary auditory cortex: There is some experimental support for the hypothesis that an auditory signal is decomposed into spectro-temporal edges that are combined into increasingly complex feature conjunctions that represent auditory events (de Charms et al., 1998; Shamma, 2001).

Amplitude and frequency transitions are important features of the auditory events and many studies have demonstrated the sensitivity of the auditory system to these transients (Eggermont, 1993; Schreiner & Langner, 1988; Phillips et al, 1988;

Heil et al, 1992; for review see Fishbach et al, 2001; Phillips et al, 2002). These studies find that a relevant parameter is the rate of change: units respond more strongly when the change is fast than slow (Fishbach et al, 2001). An ecological justification for this behavior this is that a sharp transition often implies the appearance of a new auditory object (e.g. Bregman, 1994). Fishbach et al (2001; 2003) have proposed a neural model of physiological and behavioral responses to transients that is able to account for numerous experimental findings. They interpret it as a realization of auditory temporal edge calculation.

Edge detection is an important ingredient of scene analysis and object detection, but most certainly not sufficient. The system still needs to group the edges together, to decide what object the edges correspond to, to discount spurious 'edges' due to sharp features of the object or ground, and finally where the object appeared and what one should do about it. This obviously involves further computation. As in vision, where edges are considered to be inputs into the mechanisms that construct objects (Julesz, 1971; Peterson, 1999; Geisler et al, 2001) there is evidence in the auditory domain that edges may be first stages in the computation of more complex structures. The 'gap transfer illusion' (Nakajima et al., 2001; Remijn & Nakajima, 2005) is observed when a long frequency glide with a temporal gap and a short frequency glide cross each other. At the cross point, the gap, even though physically located in the long glide, is perceived as belonging to the short frequency glide. The effect is explained by an 'event construction model' (Nakajima et al, 2004) that feeds upon the onsets and offsets detected by the system. In the model, edges behave as independent sub-events that are subject to grouping heuristics such as temporal and

spatial (spectral) proximity. It would be interesting to know whether the responses we record represent a stage before, or after, this organization. Another way to formulate this question is to ask whether onsets and offsets, as evoked responses recorded by MEG, are already bound to objects (e.g Gutschalk et al, 2004b) or whether they reflect as yet unbound events (Nakajima et al, 2004; see below).

The interpretation of the responses we observe as auditory edges is consistent with the fact that transitions between different signals (such as emergence of a tone within a noise background vs. emergence of a tone within a tone background) are manifested by activations with different characteristics (see Chapter 2): Edges are defined by the nature of the two objects between which they are located. Interpretation of edge detection as participating in object formation is in line with Kubovy and Van Valkenburg's (2001) view of auditory objects as defined by the background from which they are extracted. Kubovy and Van Valkenburg (2001) define an auditory object as anything that can be extracted from background. It is also consistent with the demonstration of extreme sensitivity of primary auditory cortical neurons to small perturbations in their acoustic input and specifically to their sensitivity to background noise as acoustic context (Bar Yosef et al, 2002).

4.3 Phenomenology of change-related MEG responses

As discussed in Chapter 1, the challenge in MEG research is to deduce the 'how' from information about 'when' and 'where'. That is, to try to understand what computations are performed and how information is being stored by observing how brain activity in response to sound unfolds over time. In the present case, such a departure from the data might be premature. On the other hand, the formulation of

concrete hypotheses about the abstract neural computations that might underlie the observed magnetic field deflections may serve to generate new predictions that are testable experimentally and in turn refine our understanding of the computational principles behind the recorded MEG signal.

In Chapter 3 we hypothesized that the M50-like peak at about 70 ms post change that is found in the first response to transitions from regularity, both in our work (CR or CC transitions) or in that of others (e.g., Ritter et al 2005, transitions between different IRNs) reflects minimal integration by a default, obligatory, cortical integration window. In contrast, the M100-like peak occurring after transitions *to* regularity, both in our work (RC or CC transitions, or also CR transitions to sequences with large pip sizes) and in that of others (e.g. Chait et al., 2006; Gutschalk et al, 2004a; Krumbholz et al., 2003, onset of a tonal object within noise, transition from noise to IRN or transition from irregular to regular click trains) is hypothesized to reflect the operation of an additional, optional, integration window. The size of this window may depend on the precise stimulus statistics, explaining the dependency of M100 latency that has been observed in our studies and those of others (dependency of latency on pip size for our RC stimuli, or on period for the stimuli of Ritter et al, 2005 or Krumbholz et al, 2003).

Another way of interpreting the functional significant of these deflections is to invoke a higher-level 'cognitive' operation of *construction* or *destruction* of a perceptual representation. What is meant here by a 'representation' is some sort of predictive model or rule. An event that violates the rule flags the occurrence of a transition. Unexpected shifts, such as our CR transitions involve detecting a mismatch

between the predictive model and the input, this operation is reflected in the M50-like peak. On the other hand, the detection of the opposite transition requires *acquiring* a rule, which is subserved by the mechanisms underlying the M100-like deflection. This is another way of describing the fundamental asymmetry we observe in our experiments.

The M100-like responses we observe to transitions, or to the onset of a tone among background noise, are similar to the omnipresent M100 onset response reported in the literature (Roberts et al., 2000; see also Seither-Preisler, 2004). The findings detailed here may therefore serve to shed light on the 'phenomenology' of the M100. The M100 onset response (or its EEG counterpart) has been hypothesized to code for diverse sound features such as phonation and articulation (Tiitinen et al, 2005) spatial cues (Palomaki et al, 2005; Soeta et al, 2005b; Soeta et al, 2005), vowel categories (Shestakova et al, 2004; Mäkelä et al, 2003) perceptual categorization of sound spectral envelopes (Mizuochi et al, 2005) as well as pitch (e.g. Pantev et al., 1996; Fujioka et al., 2003; Ragot & Lepaul-Ercole, 1996). This plethora of roles hypothesized for the M100 generating mechanism alone should raise a red flag: it cannot be that *all* these hypotheses are correct. For example, if M100 codes for vowel category, or space, its latency should not vary with pitch. As discussed in the opening Chapter (and because I do not want to be pessimistic), I believe it is more reasonable to assign a more global role to all of these deflections that is related to auditory edge detection and the signaling of the appearance of a new auditory object in the environment. Such an interpretation is consistent with the relatively long latency of the M100, its prominence, and the effects of attention and figure/ground status on the M100

response (e.g. Woldorff et al, 1993; Chait et al, 2004; Pressnitzer et al, 2004; Gutschalk et al, 2005).

Several independent lines of evidence point to the fact that the M100 response is more than a detector of the onset of energy: The data in Chapter 3 here implicate the M100 response as evoked in situations where a new representation has been constructed. The results of the tone experiment (Chapter 3.2; especially the 15ms pip size data) are particularly striking in that respect because the transitions in 'constant to random' and 'random to constant' stimuli are equally perceptible and salient yet the M100 response is only evoked in the latter case. In Chapters 1 and 2 here (see also Chait et al., 2004) I describe experiments where the M100 response is reduced when a signal is perceived as 'background' vs. situations where the task defines is as 'foreground'. Pressnitzer et al (2004) measured MEG brain responses during the continuity illusion (Bregman, 1990). They used amplitude modulated noise with easily detectable gaps imposed upon it. When the gaps are filled with loud noise, listeners report the AM as continuous. This perceptual continuity was accompanied by greatly reduced M100 onset responses after the gaps. Thus the lack of offset/onset edges may underlie the perceptually continuous auditory object.

In this context it is interesting to mention Friston (2005) who proposes that the principles underlying the generation of evoked brain responses are related to prediction error: "….the role of backward connections is to provide contextual guidance to lower levels through a prediction of the lower level's inputs. When this prediction is incomplete or incompatible with the lower areas input, a prediction error is generated that engenders changes in the area above until reconciliation" (pp 826).

Although such an interpretation is principally consistent with the MMN and with the properties of onset responses, our data suggest that although prediction error may explain the CR transitions be observe, a different kind of mechanism (one that *builds up* a prediction) is involved in generating evoked responses in the RC type transitions.

Bayesian models such as Kersten et al (2004) or Friston (2005) assign a computationally important role to feedback connections: Higher level areas modulate the responses of lower level areas by explaining away their input ("shut-up" heuristic) or by enhancing the activity of one of competing lower-level representations ("stop gossiping" heuristic). Such feedback loops are not commonly investigated in audition. In vision research, it has been shown that responses of V1 neurons are modulated by whether the input in their receptive field constitutes a perceptual figure or background (Zipser et al, 1996; Lamme, 1995). This 'contextual' modulation commences at about 40ms post onset (with a peak around 100 ms) and results from backwards connections from higher visual areas where figure/ground segregation is computed. These findings are important for audition because they suggest that the responses we observe might be generated by similar kinds of feedback architecture, the investigation of which would contribute to our understanding of the mechanisms we observe.

4.4 Early auditory cortical responses and behavior

On the one hand, I interpret the auditory cortical responses discussed in this dissertation as related to the perceptual organization of sound. At the same time, it is

clear from the comparison between MEG and behavioral responses, and especially from the analysis in Chapter 2.2, that the relationship between these early brain responses and behavior is not straightforward and that later processes affect the way listeners consciously respond to the stimuli. More experimentation is needed in order to map out the relationship between behavior and responses from auditory cortex and for understanding the perceptual role of these auditory change detectors.

Another approach to understanding the role of these mechanisms is to study subjects in whom these systems are impaired. The results of the experiment described in Chapter 2.3 suggest that dyslexics may indeed be impaired at processing such transitions. If the MEG experiments that we plan to perform with subjects from this population confirm our conjecture that this population is characterized by slow early cortical responses to change, we might have the opportunity, by comparing this pattern of responses with the spectrum of auditory impairments associated with dyslexia (Wright et al., 2000; Rosen, 2003; Ramus 2003), to tease apart the physiological and behavioral corollaries of these early, pre-attentive processes.

Acoustic environments are characterized by the irregular appearance and disappearance of sound sources. The onset of rain, the turning off of a radio, somebody beginning to speak, the opening of a door, are examples of such events. Our stimuli embodied certain characteristics of such events. A monaural RC or binaural $0 \rightarrow 1$ transition models the appearance of a regular object within an irregular background. A monaural CR or binaural $1 \rightarrow 0$ transition models the disappearance of the same object, or else the appearance of an irregular object in a regular background. The apparent symmetry of this account masks the fact that their detection involves

different requirements, evident in the longer latency of MEG responses to the former (150 ms post change) than the latter (70 ms post change). Note that this discussion is on the early cortical processing level. As discussed, our psychophysical data indicate that the relationship of this stage to behavior is not clear-cut and need to be clarified in future experiments. The brain responses we observe probably reflect an early, pre-attentive and incomplete organization that later processes operate on and bring to the attention of the listener. However the early organization of sensory input might provide clues to interpreting perceptual phenomena.

Our data may offer an explanation for the perceptual asymmetry (Bregman et al., 1994; Philips et al, 2002) between onsets and offsets. Bregman et al (1994) report that when groups of pure tones are highly overlapped in time, the ability to judge their onset order is strongly affected by the suddenness of their onsets (size of onset ramp). Abruptness of changes also had effects on the ease with which offset asynchrony was detected however the ability to detect the order of offsets was much worse than that of onsets. After the previous discussion, the asymmetry in perceptual weighting of onsets and offsets should not come as a surprise. Although they appear symmetrical when viewed in schematic form on paper (e.g. a plot of a spectrum), the one-way flow of time makes them highly asymmetrical. The appearance of symmetry is a product of our visual bias and inclination to use graphic representations to discuss these phenomena. There is an asymmetry at the neural level between onset and offset coding cells (Phillips et al, 2002), and the onset and offset responses that I observe in the MEG data also indicate that different computations accompany onsets and offsets. However, a full analysis of these contrasting responses remains to be done.

Interestingly, the MEG data predicts that offsets are detected faster than onsets in some cases, although behaviorally this is not the case. It may be that behaviorally, subject's detection is based on the mechanisms that generate the M100-like deflections in RC, CC or CR transitions and it is those processes that are weighted more strongly in detection – such an explanation would be able to account for subjects in the tone-change experiment missing more CR than RC transitions, even though theoretically the opposite pattern is predicted. Thus it may be that onset edges carry information about that an event has occurred and also contain certain information about the properties of that event whereas offset detection serves just to notify that something has ended -- Offsets are computed faster, but carry no information about *what* it is that has ended which may cause the pattern of behavior reported by Bregman et al (1994).

4.5 Beyond the auditory modality

In fact, a similar temporal and spatial morphology of onset responses is observed in all sensory modalities (e.g Lehman & Julesz, 1978; Super et al, 2001; Liu et al, 2002; Moradi et al, 2003; Heinen et al, 2005; Kekoni et al, 1992). The co-occurance of responses from different modalities at around the same time is *not* to be expected, given the very different processing requirements in different modalities. Rather, it seems more likely that these similar responses reflect operations required in all modalities such as detection of the appearance of objects. In vision, contextual modulation (the figure/ground status of a texture) has a characteristic latency of 80-100msec after stimulus onset (Zipser et al, 1996; Lamme, 1995). The properties of

this component, even for salient sensory stimuli, vary depending on the observer's attentional state, and predict whether the stimulus would be consciously detected (Super et al, 2001).

It is clear that humans exist in a multi-sensory environment, in which an object that generates sound may also have a visual form. Information from different modalities combines and interacts in complex ways in the construction of a representation of the world (e.g. Driver & Spence, 2000). This interaction may occur at a very early stage (Schroeder & Foxe, 2005) in which case there are limits as to what can be learned by studying each modality alone. Recently it has been demonstrated that patients with deficits in visual scene analysis (neglect) have corresponding deficits in auditory scene analysis (Cusack et al., 2000; 2001 see also Marcel et al, 2006), suggesting that visual and auditory scene analysis share common stages. It may be that the functional significance of these responses actually reflects a form of *binding* of incoming information across modalities (e.g. van Wassenhove et al, 2005). Such low-level multisensory binding might account for illusions such as the sound-induced visual "rabbit" (Kamitani & Shimojo, 2001; Shams et al, 2005).

Obviously further study is required to chart out the respective roles of feature detection, edge detection, event detection, object detection, etc. in the responses that we measure. Hopefully this will lead us to an understanding of the ecological significance of this activation and how it is related to behavior.

References

Ableson H, Sussman GJ, Sussman J (1996) Structure and interpretation of computer programs. Second ed. Cambridge, MA: MIT press

Adachi Y, Shimogawara M, Higuchi M, Haruta Y, Ochiai M (2001) Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. IEEE Transactions on Applied Superconductivity 11:669-672.

Adrian ED (1946) The Physical Background of Perception, Clarendon Press: Oxford,

Agnew JA, Dorn C, Eden GF (2004) Effect of intensive training on auditory processing and reading skills. Brain and Language 88:21-25.

Akeroyd MA, Moore BC, Moore GA (2001) Melody recognition using three types of dichotic-pitch stimulus. J Acoust Soc Am. 110:1498-1504.

Akeroyd MA (2001) A binaural cross-correlogram toolbox for MATLAB http://www.ihr.gla.ac.uk/ (software package) .

Akeroyd MA, Summerfield Q (1999) A binaural analog of gap detection J Acoust Soc. Am. 105:2087-2820.

Alain C, Schuler BM, McDonald KL (2002) Neural activity associated with distinguishing concurrent auditory objects. J Acoust Soc Am. 111: 990-5.

Amano K, Nishida S, Takeda T (2005): MEG responses correlated with the visual perception of speed change. Vision Research 46: 336-345

Amitay S, Ahissar M, Nelken I. (2002) Auditory processing deficits in reading disabled adults. Journal of the Association for Research in Otolaryngology. 3:302-20.

Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping. IEEE signal processing magazine. 18:14-30.

Banai K, Nicol T, Zecker SG, Kraus N. (2005) Brainstem timing: implications for cortical processing and literacy. Journal of Neuroscience. 25:9850-7.

Banai K, Ahissar M. (2004) Poor frequency discrimination probes dyslexics with particularly impaired working memory. Audiology and Neurootology. 9:328-40.

Bares M, Rektor I, Kanovsky P, Streitova H (2003) Cortical and subcortical distribution of middle and long latency auditory and visual evoked potentials in a cognitive (CNV) paradigm. Clin Neurophysiol. 114:2447-60.

Barnes J (1984) Complete works of Aristotle. Princeton, NJ: Princeton university press.

Barlow HB (1961). The coding of sensory messages. Chapter 13. In Current problems in animal behaviour, Thorpe and Zangwill (Eds), Cambridge University Press, pp. 330-360.

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. 22:8619-32.

Bendor D, Wang X (in press) The neuronal representation of pitch in primate auditory cortex. Nature

Ben-Yehudah G, Banai K, Ahissar M. (2004) Patterns of deficit in auditory temporal processing among dyslexic adults. Neuroreport. 15:627-31.

Bilsen F (1977) Pitch of noise signals: evidence for a "central spectrum". J Acoust SocAm 61:150-161.

Blauert J (1997). Spatial Hearing, revised edition. Cambridge, MA: the MIT press.

Blauert J (1972) On the lag of lateralization caused by interaural time and intensity differences. Audiology 11:265-270.

Blauert J, Lindemann W (1986) Auditory spaciousness: Some further psychoacoustic analyses, J. Acoust. Soc. Am. 80:533-541

Boehnke SE, Hall SE, Marquardt T (2002) Detection of static and dynamic changes in interaural correlation J Acoust Soc Am. 110:1617-1626.

Boring EG (1942) Sensation and perception in the history of experimental psychology. NY, NY: Appleton-Century-Crofts, Inc.

Borgmann C, Roβ B, Draganova R, Pantev C (2001) Human auditory middle latency responses: influence of stimulus type and intensity. Hear. Res. 158:57-64.

Bosnyak DJ, Eaton RA, Roberts LE (2004) Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. Cereb Cortex. 14:1088-99.

Bradley, L, Bryant P. (1983) Categorizing sounds and learning to read: A causal connection. Nature. 301:419-421.

Brazier MAB (1961) A history of the electrical activity of the brain: The first half century. London: Pitman medical publishing.

Brechmann A, Scheich H (2005) Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb Cortex. 15:578-87.

Bregman AS, Ahad PA, Kim J (1994) Resetting the pitch-analysis system. 2. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. J Acoust Soc Am. :26:94-703.

Bregman AS. (1990). Auditory Scene Analysis Cambridge, MA: MIT press.

Bretherton, L, Holmes VM (2003). The relationship between auditory temporal processing, phonemic awareness, and reading disability. Journal of Experimental Child Psychology. 84:218-243.

Brochard R, Drake C, Botte MC, McAdams S. (1999) Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation. J Exp Psychol Hum Percept Perform. 25:1742-59

Budd TW, Hall DA, Goncalves MS, Akeroyd MA, Foster JR, Palmer AR, Head K, Summerfield AQ (2003) Binaural specialisation in human auditory cortex: an fMRI investigation of interaural correlation sensitivity. Neuroimage. 20:1783-94.

Budd TW, Barry RJ, Gordon E, Rennie C, Michie PT (1998) Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness Int J Psychophysiol. 31:51-68.

Burr DC, Morrone MC, Spinelli D (1989) Evidence for edge and bar detectors in human vision. Vision Res. 1989;29(4):419-31.

Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol. 76:1698-1716.

Carlyon RP (2004) How the brain separates sounds. Trends Cogn Sci. 8:465-71.

Carlyon RP, Demany L, Deeks J (2001) Temporal pitch perception and the binaural system. J Acoust Soc Am. 109:686-700.

Carpenter RHS (2004) Contrast, probability, and saccadic latency: Evidence for independence of detection and decision. Curr. Biol. 14:1576-80.

Carpenter RHS (1999a) A neural mechanism that randomizes behavior. J. Consciousness studies 6:13-22.

Carpenter RHS (1999b) Visual selection: Neurons that make up their minds. Curr. Biol 9:R595-R598.

Carr CE (1993) Processing of temporal information in the brain. Annu. Rev. Neurosci. 16 :223-243.

Cattel JM (1886) The time taken up by cerebral operations. Mind 11:220:242.

Chait M, Simon JZ (in press) The dynamics of the construction of auditory perceptual representations. In Reasoning and Cognition Interdisciplinary Series on Reasoning Studies, Vol.2 (Andler et al (eds)). Tokyo, Japan: Keio University Press.

Chait M, Poeppel D, Simon JZ. (epub 2005) Neural Response Correlates of Detection of Monaurally and Binaurally Created Pitches in Humans. Cereb Cortex.

Chait M, Poeppel D, de Cheveigne' A, Simon JZ. (2005) Human Auditory cortical processing of changes in interaural correlation. J Neurosci.25:8518 –27

Chait M, Simon JZ, Poeppel D. (2004) Auditory M50 and M100 responses to broadband noise: functional implications. Neuroreport. 15:2455-8.

Cherry EC (1953). Some experiments on the recognition of speech, with one and with

two ears. Journal of the acoustical society of America. 25:975-979.

Colburn HS (1995).Computational models of binaural processing. In: Auditory
computation - Springer handbook of auditory research: v. 6. (Hawkins HL et
al. (eds)), pp. 332-400 New York: Springer-Verlag.

Colburn HS, Durlach NI (1978) Models of binaural interaction. . In: Handbook of
perception, Vol IV (Carterette EC, Friedman, M, eds.), pp. 467-518. NY:
Academic press.

Cramer E, Huggins WH (1958) Creation of pitch through binaural interaction. J
Acoust Soc Am 30:413-417.

Culling JF, Colburn HS Spurchise M (2001) Interaural correlation sensitivity J
Acoust Soc Am. 110:1020-1029.

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

Culling J, Marshall D, Summerfield Q (1998) Dichotic pitches as illusions of binaural
masking release I: Huggins' pitch and the binaural edge pitch" J. Acoust. Soc.
Am. 103**:**3509-3526.

Cusack R, Carlyon RP (2003) Perceptual asymmetries in audition. J Exp Psychol
Hum Percept Perform. 29:713-25.

Cusack R, Carlyon RP, Robertson IH (2001) Auditory midline and spatial
discrimination in patients with unilateral neglect. Cortex. 37:706-9.

Cusack R, Carlyon RP, Robertson IH (2000) Neglect between but not within auditory
objects. J Cogn Neurosci. 12:1056-65.

Dean I, Harper NS, McAlpine D (2005) Neural population coding of sound level

adapts to stimulus statistics. Nat Neurosci. 8:1684-9.

deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for

cortical neurons. Science. 280:1439-43.

de Cheveigne, A (2005) Pitch processing models. In: Pitch: Neural coding and

perception (Plack CJ, Oxenham, AJ, Fay RR, Popper AN (eds)), pp. 169-233.

NY, NY: Springer.

de Cheveigné, A. (1993). "Separation of concurrent harmonic sounds: Fundamental

frequency estimation and a time-domain cancellation model of auditory

processing." J. Acoust. Soc. Am. 93: 3271-3290.

Demany L, Ramos C. (2005) On the binding of successive sounds: perceiving shifts

in nonperceived pitches. J Acoust Soc Am. 117:833-41.

Denckla M, Rudel RG (1976). Rapid automatized naming (RAN): Dyslexia

differentiated from other learning disabilities. Neuropsychologia, 14, 471-479.

Devlin JT, Raley J, Tunbridge E, Lanary K, Floyer-Lea A, Narain C, Cohen I,

Behrens T, Jezzard P, Matthews PM, Moore DR (2003) Functional

asymmetry for auditory processing in human primary auditory cortex. J

Neurosci 23:11516-11522.

DeWeese M., Zador A. (1998) Asymmetric dynamics in optimal variance

adaptation, Neural Computation, 10:1179-1202 .

Donders FC (1868) La vitesse des actes psychiques. Archives Neerlandaises 3 :269-

317.

Dougherty RF, Cynader MS, Bjornson BH, Edgell D, Giaschi DE. (1998) Dichotic
pitch a new stimulus distinguishes normal and dyslexic auditory function.
Neuroreport. 9:3001-5.

Driver J, Spence C (2000) Multisensory perception: beyond modularity and
convergence. Curr Biol. 10:R731-5.

Dupoux E. (1993) The time course of prelexical processing: The syllabic hypothesis
revisited. G. Altmann & R. Shillcock (Eds.) Cognitive Models of Speech
Processing (pp. 81-114). Hillsdale, NJ: Erlbaum.

Durlach NI, Gabriel KJ, Colburn HS, Trahiotis C (1986) Interaural correlation
discrimination: II. Relation to binaural unmasking J Acoust Soc Am.
79:1548-57.

Durlach NI (1962) Note on the creation of pitch through binaural interaction J Acoust
Soc Am 34:1096-1099.

Durlach NI (1963) Equalization and cancellation theory of binaural masking-level
differences J Acoust Soc Am 35:1206-1218.

Dyson BJ, Alain C, He Y (2005) I've heard it all before: perceptual invariance
represented by early cortical auditory-evoked responses. Brain Res Cogn
Brain Res.   23:457-60.

Dyson BJ, Alain C (2004) Representation of concurrent acoustic objects in primary
auditory cortex J Acoust Soc Am. 115:280-8.

Eden GF, Zeffiro TA. (1998) Neural systems affected in developmental dyslexia
revealed by functional neuroimaging. Neuron. 21:279-82.

Eden GF, VanMeter JW, Rumsey JM, Maisog JM, Woods RP, Zeffiro TA. (1996)

Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. Nature. 382:66-9.

Efron B, Tibshirani RJ (1993) An introduction to the Bootstrap. NY: Chapman and Hall.

Eggermont JJ (1993) Differential effects of age on click-rate and amplitude modulation- Frequency coding in primary auditory cortex of the cat. Hear Res. :175-92.

Eramudugolla R, Irvine DR, McAnally KI, Martin RL, Mattingley JB. (2005) Directed attention eliminates 'change deafness' in complex auditory scenes. Curr Biol. 15:1108-13.

Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. Nature. 412:787-92.

Fishbach A, Yeshurun Y, Nelken I (2003) Neural model for physiological responses to frequency and amplitude transitions uncovers topographical order in the auditory cortex. J Neurophysiol. 90:3663-78.

Fishbach A, Nelken I, Yeshurun Y (2001) Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. J Neurophysiol. 85:2303-23.

Fletcher H (1940) Auditory patterns. Rev. Mod. Phys. 12, 47–65.

Fodor JA. (1983) The Modularity of Mind, Cambridge, MA: MIT Press.

Friederici AD (2005) Neurophysiological markers of early language acquisition: from syllables to sentences. Trends Cogn Sci. 9:481-8.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci.

360:815-36.

Fujioka T, Ross B, Okamoto H, Takeshima Y, Kakigi R, Pantev C (2003) Tonotopic representation of missing fundamental complex sounds in the human auditory cortex. Eur J Neurosci. 18:432-40.

Gabriel KJ, Colburn HS (1981) Interaural correlation discrimination: I. Bandwidth and level dependence J Acoust Soc Am. 69:1394-1401.

Gage N, Roberts TP, Hickok G (2006) Temporal resolution properties of human auditory cortex: Reflections in the neuromagnetic auditory evoked M100 component. Brain Res. 1069:166-71.

Gage NM, Roberts TP (2000) Temporal integration: reflections in the M100 of the auditory evoked field. Neuroreport. 11:2723-6.

Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N. (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. Annals of Neurology. 18:222-33.

Galaburda AM, Menard MT, Rosen GD (1994) Evidence for aberrant auditory anatomy in developmental dyslexia. Proceedings of the National Academy of Sciences of the U S A. 91:8010-3.

Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. Vision Res. 41:711-24.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liegeois-Chauvel C (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin Neurophysiol. 112:1850-9.

Gold, JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. Trends Cogn Sci. 5:10-16

Goldstein EB (2005) Blackwell handbook of sensation and perception. Malden, MA: Blackwell publishing.

Grantham (1995) Spatial Hearing and related phenomena. In: Hearing: handbook of perception and cognition, 2nd ed (Moore BCJ, ed.), pp. 297-346. San Diego: Academic Press

Grantham DW (1982) Detectability of time-varying interaural correlation in narrow-band noise stimuli. J Acoust Soc Am. 72:1178-84.

Grantham DW, Wightman FL (1979) Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation J. Acoust. Soc. Am. 65:1509–1517.

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

Green DM (1993) Auditory intensity discrimination. In: Human psychophysics (Yost WA, Popper AN, Fay RR, (eds)), pp. 13-55. . NY, NY: Springer-Verlag.

Greenberg S, Poeppel D, Roberts T (1998) A space-time Theory of pitch and timbre based on cortical expansion of the cochlear traveling wave delay. In: Psychophysical and physiological advances in hearing (Palmer AR et al, eds.), pp. 293-300. London: Whurr Publishers.

Greenhouse SW, Geisser S (1959) On methods in the analysis of profile data. Psychometrika 24:95-112.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci. 5:887-92.

Griffiths TD, Büchel C, Frackowiak RSJ, Patterson RD (1998a) Analysis of temporal structure in sound by the human brain. Nat. Neurosci. 1:422-427

Griffiths TD, Rees G, Rees D, Green GGR, Witton C, Rowe D, Büchel C, Turner R, Frackowiak RSJ (1998b) Right parietal cortex is involved in the perception of sound movement in humans. Nat. Neurscience 1:74-79

GriffithsTD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD (2001) Encoding of the temporal regularity of sound in the human brainstem. Nat Neurosci. 4:633-637

Grigorenko EL, Wood FB, Meyer MS, Pauls JED, Hart LA, Pauls DL (2001). Linkage studies suggest a possible locus for developmental dyslexia on chromosome 1p. American Journal of Medical Genetics (Neuropsychiatric Genetics) 105: 120-129.

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

Gruber WR, Klimesch W, Sauseng P, Doppelmayr M. (2005) Alpha phase synchronization predicts P1 and N1 latency and amplitude size. Cereb Cortex. 15:371-7.

Gueguin M, Le Bouquin-Jeannes R, Faucon G, Chauvel P, Liegeois-Chauvel C (2006) Evidence of Functional Connectivity between Auditory Cortical Areas Revealed by Amplitude Modulation Sound Processing. Cereb Cortex. [Epub ahead of print]

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ (2005) Neuromagnetic correlates of streaming in human auditory cortex. J Neurosci. 25:5382-8.

Gutschalk A, Patterson RD, Scherg M, Uppenkamp S, Rupp A (2004a) Temporal dynamics of pitch in human auditory cortex. Neuroimage. 22:755-66.

Gutschalk A, Patterson RD, Uppenkamp S, Scherg M, Rupp A (2004b) Recovery and refractoriness of auditory evoked fields after gaps in click trains. Eur J Neurosci. 20:3141-7.

Gutschalk A, Patterson RD, Rupp A, Uppenkamp S, Scherg M (2002) Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage. 15:207-16.

Hairston WD, Burdette JH, Flowers DL, Wood FB, Wallace MT. (2005) Altered temporal profile of visual-auditory multisensory interactions in dyslexia. Experimental Brain Research. 166474-80.

Hall DA, Barrett DJ, Akeroyd MA, Summerfield AQ (2005) Cortical representations of Temporal structure in sound. J Neurophysiol. 94:3181-91.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounamaa OV (1993) Magnetoencephalography- theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys., 65:413-497.

Hanes DP, Schall JD (1996) Neural control of voluntary movement initiation. Science 274:427-430.

Hautus MJ, Johnson BW (2004) Object-related brain potentials associated with the

perceptual segregation of dichotically embedded pitch. . J Acoust Soc Am 117:275-280.

Hari R, Renvall H (2001). Impaired processing of rapid stimulus sequences in dyslexia.Trends in cognitive sciences. 5:525-532.

Hari R, Levanen S, Raij T (2000) Timing of human cortical functions during cognition: role of MEG. Trends Cogn Sci. 4:455-462.

Hari R (1990) The neuromagnetic method in the study of the human auditory cortex. In: Auditory evoked magnetic fields and the electric potentials (Grandori, F. et al (eds)), pp 222-282 Basel: Krager-Verlag.

Heil P, Langner G, Scheich H (1992) Processing of frequency-modulated stimuli in the chick auditory cortex analogue: evidence for topographic representations and possible mechanisms of rate and directional sensitivity. J Comp Physiol [A]. 171:583-600.

Heinen K, Jolij J, Lamme VA (2005) Figure-ground segregation requires two distinct periods of activity in V1: a transcranial magnetic stimulation study. Neuroreport. 16:1483-7.

Helenius P, Uutela K, Hari R. (1999) Auditory stream segregation in dyslexic adults. Brain. 122:907-13.

Hertrich I, Mathiak K, Menning H, Lutzenberger W, Ackermann H (2005) MEG responses to rippled noise and Huggins pitch reveal similar cortical representations. Neuroreport 16:193-196.

Hill NI, Bailey PJ, Griffiths YM, Snowling MJ (1998) Frequency acuity and binaural

masking release in dyslexic listeners. Journal of the Acoustical Society of America . 106:L53-8.

Hillyard SA, Picton TW (1978) On and off components in the auditory evoked potential. Percept Psychophys. 24:391-8.

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

Houtsma AJM, Goldstein JL (1971) The central origin of the pitch of complex tones: evidence from musical interval recognition. J Acoust Soc Am 51:520-529.

Hubel DH, Wiesel TN. (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat.. J Neurophysiol. 28:229-89.

Jancke L, Mirzazade S, Shah NJ (1999) Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. Neurosci Lett. 266:125-8.

Jeffress LA (1948) A place theory of sound localization. J. Comp. Physiol. Psychol., 41:35-39.

Jiang D, McAlpine D, Palmer AR (1997a) Responses of neurons in the inferior colliculus to binaural masking level difference stimuli measured by rate-versus-level functions. J. Neurophysiol. 77:3085-3106.

Jiang D, McAlpine D Palmer AR (1997b) Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. J. Neurophysiol. 17:9331–9339.

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

:2245-50.

Johnsrude IS, Penhune VB, Zatorre RJ (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123:155-163.

Jones SJ (2002a)The internal auditory clock: what can evoked potentials reveal about the analysis of temporal sound patterns, and abnormal states of consciousness? Neurophysiol Clin. 32:241-53.

Jones SJ, Sprague L, Vaz Pato M (2002b) Electrophysiological evidence for a defect in the processing of temporal sound patterns in multiple sclerosis. J Neurol Neurosurg Psychiatry. 73:561-7.

Jones SJ, Perez N (2002) The auditory C-process of spectral profile analysis. Clin Neurophysiol. 113:1558-65.

Jones SJ, Pitman JR, Halliday AM (1991) Scalp potentials following sudden coherence and discoherence of binaural noise and change in the inter-aural time difference: a specific binaural evoked potential or a "mismatch" response? Electroencephalogr Clin Neurophysiol. 80:146-54.

Joris PX, van de Sande B, Recio-Spinoso A, van der Heijden M (2006) Auditory midbrain and nerve responses to sinusoidal variations in interaural correlation. J Neurosci. 26:279-89.

Joris PX, Smith PH, Yin TC (1998) Coincidence detection in the auditory system: 50 years after Jeffress. Neuron. 21: 1235-8.

Julesz B, Tyler CW. (1976) Neurontropy, an entropy-like measure of neural correlation, in binocular fusion and rivalry Biol Cybern. 1976 23:25-32.

Julesz B (1971) Foundations of Cyclopean Perception. Chicago, IL: The University

of Chicago Press.

Julesz B (1962) Visual pattern discrimination. IRE Trans. Inf. Theory 8:84:92

Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. Curr. Opin Neurobiol. 9:164-170.

Kamitani Y, Shimojo S (2001) Sound-induced visual "rabbit". Journal of vision 1(3): 478a

KaufmanL, Lu ZL (2003) Basic Neuromagnetism and Magnetic source imaging. In: magnetic source imaging of the human brain. (Lu ZL, Kaufman L (eds)) pp 1-42. Mahwah NJ: Lawrence Erlbaum

Kekoni J, Tiihonen J, Hamalainen H (1992) Fast decrement with stimulus repetition in ERPs generated by neuronal systems involving somatosensory SI and SII cortices: electric and magnetic evoked response recordings in humans. Int J Psychophysiol. 12:281-8.

Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Annu Rev Psychol. 55:271-304.

Klein MA, Hartmann WM (1981) Binaural Edge Pitch J Acoust Soc Am 70:51-59.

Klimesch W, Schack B, Schabus M, Doppelmayr M, Gruber W, Sauseng P (2004) Phase-locked alpha and theta oscillations generate the P1-N1 complex and are related to memory performance. Brain Res Cogn Brain Res. 19:302-16.

Kollmeier B, Gilkey RH (1990) Binaural forward and backward masking: Evidence for sluggishness in binaural detection. J Acoust Soc Am 87:1709-1719.

Krumbholz K, Schonwiesner M, Rubsamen R, Zilles K, Fink GR, von Cramon DY (2005) Hierarchical processing of sound location and motion in the human

brainstem and planum temporale. Eur J Neurosci. 21:230-8.

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lütkenhöner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex 13:765-772.

Kubovy M, Van Valkenburg D. (2001) Auditory and visual objects. Cognition. 80:97-126.

Kubovy M (1981) Concurrent pitch segregation and the theory of indispensable attributes. In Perceptual organization (Kubovy M, Pomerantz JR (eds)). Hillsdale, NJ: Lawrence Erlbaum publishers.

Kujala A, Naatanen R (2003) Auditory environment and change detection as indexed by the mismatch negativity (MMN). In Polich, J (ed): Detection of change: event related potential and fMRI findings. Boston, USA: Kluwer academic publishers. Pp: 1:22

Kujala T, Belitz S, Tervaniemi M, Naatanen R.(1999) Auditory sensory memory disorder in dyslexic adults as indexed by the mismatch negativity. European Journal of Neuroscience 17:1323-7.

Kujala T, Myllyviita K, Tervaniemi M, Alho K, Kallio J, Naatanen R. (2000) Basic auditory dysfunction in dyslexia as demonstrated by brain activity measurements. Psychophysiology. 37:262-6.

Kutas M, Federmeier KD (2000). Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn Sci. 4:463-470.

Kutas M, McCarthy G, Donchin E (1977) Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time Science 197:792-795.

Lamme VA, Rodriguez-Rodriguez V, Spekreijse H. (1999) Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the   macaque monkey. Cereb Cortex. 1999 Jun;9(4):406-13.

Lamme VA (1995) The neurophysiology of figure-ground segregation in primary visual cortex. J Neurosci. 15:1605-15.

Lehmann D, Julesz B (1978) Lateralized cortical potentials evoked in humans by dynamic random-dot stereograms. Vision Res. 18:1265-71.

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

Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P. (2004) Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex. 14:731-40.

Lindamood PC, Lindamood P (2004). Lindamood Auditory Conceptualization Test – Third Edition. Austin, TX: Pro-Ed.

Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: an MEG study. Nat Neurosci. 5:910-6.

Livingstone MS, Rosen GD, Drislane FW, Galaburda AM. (1991) Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. Proceedings of the National Academy of Sciences of the U S A. 88:7943-7.

Lu ZL, Kaufman L (2003) Magnetic source imaging of the human brain. Mahwah, NJ: Lawrence Erlbaum publishers.

Lu ZL, Dohser BA (1998) External noise distinguishes attention mechanisms. Vision Research, 38:1172-1198.

Luce, RD (1986) Response Times: Their role in inferring elementary mental
organization, NY: Oxford University Press.

Lütkenhöner B, Steinsträter O (1998) High-precision neuromagnetic study of the
functional organization of the human auditory cortex. Audiol. Neurootol
3:191-213.

Lütkenhöner B, Lammertmann C, Knecht S (2001) Latency of auditory evoked field
deflection N100m ruled by pitch or spectrum? Audiol Neurootol. 6:263-78.

Lütkenhoner B, Krumbholz K, Lammertmann C, Seither-Preisler A, Steinstrater O,
Patterson RD (2003) Localization of primary auditory cortex in humans by
magnetoencephalography. Neuroimage. 18:58-66.

Mäkelä AM, Alku P, Tiitinen H (2003) The auditory N1m reveals the left-
hemispheric representation of vowel identity in humans. Neurosci Lett.
353:111-4.

Marcel A, Mackintosh B, Postma P, Cusack R, Vuckovich J, Nimmo-Smith I, Cox
SM (2006) Is susceptibility to perceptual migration and fusion modality-
specific or multimodal? Neuropsychologia. 44:693-710.

Marr D. (1982). Vision. San Francisco, CA.:W. H. Freeman

Marr D, Hildreth E (1980) Theory of edge detection. Proc R Soc Lond B Biol Sci.
207:187-217.

Marshall CM, Snowling MJ, Bailey PJ (2001) Rapid auditory processing and
phonological ability in normal readers and readers with dyslexia. Journal of
Speech Language and Hearing Research. 44:925-40.

Martin BA, Boothroyd A (2000) Cortical, auditory, evoked potentials in response to

changes of spectrum and amplitude. J Acoust Soc Am. 107:2155-61.

Maye J, Werker JF, Gerken L (2002) Infant sensitivity to distributional information can affect phonetic discrimination. Cognition. 82:B101-11.

Mazaheri A, Jensen O (2006) Posterior {alpha} activity is not phase-reset by visual stimuli. Proc Natl Acad Sci U S A. ; [Epub ahead of print]

McAlpine D, Grothe B (2003) Sound localization and delay lines--do mammals fit the model?. Trends Neurosci. 26 :347-50.

McAnally KI, Stein JF (1996) Abnormal auditory transient brainstem function in dyslexia. Proceedings of the Royal Society B., 263: 961-965.

McCarthy G, Donchin E (1981) A metric of though: A comparison of P300 latency and reaction time. Science 211:77-80

McEvoy L, Hari R, Imada T, Sams M (1993) Human auditory cortical mechanisms of sound lateralization: II. Interaural time differences at sound onset. Hear Res. 67:98-109.

McFadden D (1975) Masking and the binaural system. In: The nervous system Vol 3: Human communication and it's disorders (Tower DB, ed), pp. 137-146. NY: Raven Press.

Meng H, Smith SD, Hager KM, Held M, Liu J, Olson RK, Penningtom BF et al. (2005) DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proceedings of the National Academy of Sciences of the U S A. 102:17053-58.

Micheyl C, Carlyon RP, Cusack R, Moore BCJ (2005) Performance measures of

perceptual auditory organization. In: Auditory signal processing: physiology, psychoacoustics, and models (Pressnitzer D., de Cheveigné A., McAdams S. Collet L., (eds)). pp: 485-494. New York: Springer Verlag.

Mizuochi T, Yumoto M, Karino S, Itoh K, Yamakawa K, Kaga K (2005) Perceptual categorization of sound spectral envelopes reflected in auditory-evoked N1m. Neuroreport. 16:555-8.

Moore BCJ (1993) Frequency analysis and pitch perception. In: Human psychophysics (Yost WA, Popper AN, Fay RR, (eds)), pp. 56-115. . NY, NY: Springer-Verlag.

Moore BCJ (1997) An introduction to the psychology of hearing. 4th ed. London: Academic press.

Moore BCJ, Glasberg BR (1983) Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. J. Acoust. Soc. Am. 74:750-753.

Moradi F, Liu LC, Cheng K, Waggoner RA, Tanaka K, Ioannides AA (2003) Consistent and precise localization of brain activity in human primary visual cortex by MEG and fMRI. Neuroimage. 18:595-609.

Naatanen R (1992) Attention and brain function. Hillsdale, NJ: Lawrence Erlbaum Associates.

Naatanen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24:375-425.

Naatanen R (1985) Selective attention in auditory information processing as revealed by event related potentials, magnetoencephalogram and regional cerebral

blood flow. In MI Posner & OSM Marin (Eds.), Attention and performance XI (pp.355-373). Hillsdale, NJ: Erlbaum.

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich MM. (1999) Cortical auditory signal processing in poor readers. Proceedings of the National Academy of Sciences of the U S A. 96:6483-8.

Nakajima Y, Sasaki T, Remijn GB, Ueda K (2004) Perceptual organization of onsets and offsets of sounds. J Physiol Anthropol Appl Human Sci. 23:345-9.

Nakajima Y, Sasaki T, Kanafuka K, Miyamoto A, Remijn G, ten Hoopen G (2000) Illusory recouplings of onsets and terminations of glide tone components. Percept Psychophys. 62:1413-25

National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and it's implications for reading instruction. National institute of child health and human development, Washington DC.

Neelon MF, Williams J, Garell PC (2006) The effects of auditory attention measured from human electrocorticograms. Clin Neurophysiol. 117:504-21.

Nelken I. (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol. 14:474-80.

Nicholson RI, Fawcett AJ (1994). Reaction times and dyslexia. The quarterly journal of Experimental psychology. 47A:29-48.

Noda K, Tonoike M, Doi K, Koizuka I, Yamaguchi M, Seo R, Matsumoto N, Noiri T, Takeda N, Kubo T (1998) Auditory evoked off-response: its source distribution is different from that of on-response. Neuroreport. 9:2621-5.

Nunez PL, Silberstein RB (2000) On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense? Brain Topogr. 13:79-96.

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia, 9:97-113.

Oxenham AJ, Bernstein JG, Penagos H. (2004) Correct tonotopic representation is necessary for complex pitch perception Proc Natl Acad Sci U S A. 101:1421-5.

Palmer AR, Jiang D, McAlpine D (2000) Neural responses in the Inferior colliculus to binaural masking level differences created by inverting the noise in one ear. J. Neurophysiol. 84:844-852.

Palmer AR, Jiang D, McAlpine D (1999) Desynchronizing responses to correlated noise: A mechanism for binaural masking level differences at the inferior colliculus J Neurophysiol. 81:722-34.

Palomaki KJ, Tiitinen H, Makinen V, May PJ, Alku P (2005) Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques.   Brain Res Cogn Brain Res. 24:364-79.

Pantev C, Hoke M, Lütkenhöner B, Lehnertz K (1989) Tonotopic organization of the auditory cortex: Pitch versus frequency representation. Science 246:486-488.

Pantev C, Eulitz C, Hampson S, Ross B, Roberts LE (1996b) The auditory evoked "off" response: sources and comparison with the "on" and the "sustained" responses. Ear Hear. 17:255-65.

Pantev C, Elbert T, Ross B, Eulitz C, Terhardt E (1996a) Binaural fusion and the

representation of virtual pitch in the human auditory cortex. Hear Res. 100:164-70.

Pantev C, Bertrand O, Eulitz C, Verkindt C, Hampson S, Schuierer G, Elbert T (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. Electroencephalogr Clin Neurophysiol. 94:26-40.

Pantev C, Ross B, Fujioka T, Trainor LJ, Schulte M, Schulz M (2003) Music and learning-induced cortical plasticity. Ann N Y Acad Sci. 999:438-50.

Papanicolaou AC, Panagiotis SG, Breier JI, Zouridakis G, Willmore J, Wheless JW, Constantinou JEC, Maggio W, Gormley, WB (1999) Magnetoencephalographic mapping of the language-specific cortex. J. Neurosurg. 90:85-93.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767-776.

Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci. 24:6810-5.

Penny WD, Kiebel SJ, Kilner JM, Rugg MD (2002) Event-related brain dynamics. Trends Neurosci. 25:387-9.

Peterson M (1999) Organization, segregation and object recognition. Intellectica 28:37-51.

Petkov CI, O'connor KN, Benmoshe G, Baynes K, Sutter ML. (2005) Auditory

perceptual grouping and attention in dyslexia. Cognitive Brain Research. 24:343-54.

Phillips DP, Hall SE, Boehnke SE (2002) Central auditory onset responses, and temporal asymmetries in auditory perception. Hear Res. 167:192-205.

Phillips DP (1988) Effect of tone-pulse rise time on rate-level functions of cat auditory cortex neurons: excitatory and inhibitory processes shaping responses to tone onset. J Neurophysiol. 59:1524-39.

Picton TW, Alain C, Woods DL, John MS, Scherg M, Valdes-Sosa P, Bosch-Bayard J, Trujillo NJ (1999) Intracerebral sources of human auditory-evoked potentials. Audiol Neurootol. 4:64-79.

Plack CJ, Oxenham A, Fay RR, Popper AN (2005). Pitch - Neural coding and Perception. NY: Springer.

Plack CJ, Moore BCJ (1990) Temporal window shape as a function of frequency and level. J. Acoust. Soc. Am. 87: 2178-2187.

Poeppel D (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication 41: 245-255.

Poeppel D, Yellin E, Phillips C, Roberts TP, Rowley HA, Wexler K, Marantz A (1996) Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds Brain Res Cogn Brain Res. 4:231-42.

Polich J (2003) Detection of change: event related potential and fMRI findings. Boston, MA: Kluwer academic press.

Pollack I, Trittipoe WJ (1959a) Binaural listening and interaural noise cross correlation J. Acoust. Soc. Am. 31: 1250–1252.

Pollack I, Trittipoe WJ (1959b) Interaural noise correlations: Examination of variables J. Acoust. Soc. Am. 31: 1616–1618.

Polyakov A, Pratt H (1998) The effect of binaural masking noise disparity on human auditory brainstem binaural interaction components. Audiology 37:17-26.

Pouliot M, Grondin S (2005) A Response-Time Approach for Estimating Sensitivity to Auditory Tempo Changes. Music Perception 22:389-399.

Pressnitzer D, Hupé JM (2005) Is auditory streaming a bistable percept? Proceedings of Forum Acusticum, Budapest. pp. 1557-1561.

Pressnitzer D, Ragot R, Ducorps A, Schwartz D, Baillet S. (2004). Is the continuity illusion based on a change-detection mechanism? A MEG study. In Joint Congress on Acoustics CFA/DAFA'04. pp. 589-590: Strasbourg.

Pylkkanen L, Feintuch S, Hopkins E, Marantz A (2004) Neural correlates of the effects of morphological family frequency and family size: an MEG study. Cognition. 91:B35-45.

Raatgever J, Bilsen F (1986) A central spectrum theory of binaural processing-Evidence from dichotic pitch. J Acoust Soc Am 80:429-441.

Ragot R, Lepaul-Ercole R (1996) Brain potentials as objective indexes of auditory pitch extraction from harmonics. NeuroReport 7:905-909.

Ramus F. (2003) Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? Current Opinion in Neurobiology. 13:212-8.

Ramus F. (2004) Neurobiology of dyslexia: a reinterpretation of the data. Trends in

Neurosciences. 27:720-6.

Reddi BAJ, Asrress KN, Carpenter RHS (2003) Accuracy, information and response time in a saccadic decision task. J Neurophys. 90:3538-3546.

Reddie BAJ, Carpetner RHS (2000) The influence of urgency on decision time. Nat Neurosci 3:827-830.

Remez RE, Rubin PE, Berns SM, Pardo JS, Lang JM (1994). On the perceptual organization of speech. Psychological Review, 101: 129-156

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. 31:183-92.

Richardson E, DiBenedetto B. (1985). Decoding Skills Test. Parkton, MD: York Press.

Ritter S, Gunter Dosch H, Specht HJ, Rupp A (2005) Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. Neuroimage 27:533-

Roberts T, Poeppel D (1996) Latency of auditory evoked M100 as a function of tone frequency NeuroReport 7:1138-1140.

Roberts T, Ferrari P, Stufflebeam S, Poeppel D (2000) Latency of the auditory evoked neuromagnetic field components: Stimulus dependence and insights towards perception. J Clin Neurophysio17:114-129.

Robertson LC (2003) Binding, spatial attention and perceptual awareness. Nat. Rev. Neurosci 4:93-102

Rosen S (2003). Auditory processing in dyslexia and specific language impairment: is there a deficit? What is it's nature? Does it explain anything? Journal of Phonetics 31:509-527

Rosner J (1979). Helping Children Overcome Learning Difficulties. (2nd ed.) New York, NY: Walker and Company.

Ross B, Herdman AT, Wollbrink A, Pantev C (2004) Auditory cortex responses to the transition from monophonic to pseudo-stereo sound. Neurol Clin Neurophysiol. 2004:18.

Roth WT, Krainz SK, Ford JM, Tinkelberg JR, Rothbart RM, Kopell BS (1976) Parameters of temporal recovery of the human auditory evoked potential. Electroencephalogr Clin Neurophysiol. 40:623-632.

Rupp A, Uppenkamp S, Gutschalk A, Beucker R, Patterson RD, Dau T, Scherg M. (2002) The representation of peripheral neural activity in the middle-latency evoked field of primary auditory cortex in humans(1). Hear Res. 174:19-31.

Sams M, Hari R, Rif J, Knuutila J (1993) The human auditory sensory memory trace persists about 10s: Neuromagnetic evidence. J. Cog. Neurosci. 5:363-370.

Saffran JR, Senghas A, Trueswell JC (2001) The acquisition of language by children. Proc Natl Acad Sci U S A. 98:12874-5.

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition. 70:27-52.

Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. Science. 274:1926-8.

Sato T, Murthy A, Thomson KG, Schall JD (2001) Search efficiency but not response

interference affects visual selection in frontal eye field. Neuron 30:1-20.

Salajegheh A, Link A, Elster C, Burghoff M, Sander T, Trahms L, Poeppel D (2004) Systematic latency variation of the auditory evoked M100: from average to single-trial data. Neuroimage. 23:288-95.

Scheich H, Baumgart F, Gaschler-Markefski B, Tegeler C, Tempelmann C, Heinze HJ, Schindler F, Stiller D (1998) Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. Eur J Neurosci. 10:803-9.

Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J Neurophysiol. 60:1823-40.

Scherg M, Von Cramon D (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr Clin Neurophysiol. 62:32-44.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory' processing. Curr Opin Neurobiol. 15:454-8.

Schulte-Korne G, Deimel W, Bartling J, Remschmidt H. (1999) Pre-attentive processing of auditory patterns in dyslexic human subjects. Neuroscience Letters. 276:41-4.

Scott SK (2005) Auditory processing--speech, space and auditory objects. Curr Opin Neurobiol. 15:197-201.

Schwartz R (2004) Perception. Malden, MA: Blackwell Publishing.

Seither-Preisler A, Patterson R, Krumbholz K, Seither S, Lütkenhoner B (2006) Evidence of pitch processing in the N100m component of the auditory evoked field. Hear Res.; [Epub ahead of print]

Seither-Preisler A, Krumbholz K, Patterson R, Seither S, Lütkenhoner B (2004) Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generators. Eur J Neurosci. 19:3073-80.

Sereno SC, Rayner K (2003) Measuring word recognition in reading: eye movements and event-related potentials. Trends Cogn Sci. 7:489-93.

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci. 23:5545-52.

Shamma S (2001) On the role of space and time in auditory processing. Trends Cogn Sci.5:340-348.

Shams L, Iwaki S, Chawla A, Bhattacharya J (2005) Early modulation of visual cortex by sound: an MEG study. Neurosci Lett. 378:76-81.

Shaywitz S, Shaywitz B, Blachman K, Pugh R, Fulbright P, Skudlarski et al. (2004) Development of left occipito-temporal systems for skilled reading in children after a phonologically-based intervention Biol Psychiatry 55: 926-933.

Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M (2004). Orderly cortical representation of vowel categories presented by multiple exemplars. Brain Res Cogn Brain Res. 21:342-50.

Shinn-Cunningham B, Kawakyu K (2003) Neural representation of cource direction in reverberant space. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics October 19-22, 2003, New Paltz, NY

Shtyrov Y, Kujala T, Lyytinen H, Ilmoniemi RJ, Näätänen R (2000) Auditory cortex evoked magnetic fields and lateralization of speech processing. NeuroReport 11:2893-2896.

Simos PG, Fletcher JM, Bergman E, Breier JI, Foorman BR, Castillo EM, Davis RN, Fitzgerald M, Papanicolaou AC (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. 58:1203-13.

Snowling M, Goulandris N, Bowlby M, Howell P (1986) Segmentation and speech Perception in relation to reading skill: a developmental analysis. Journal of Experimental Child Psychology 41:489-507.

Snowling M (2000) Dyslexia 2-nd edition. Oxford, UK: Blackwell publishers.

Soeta Y, Nakagawa S, Tonoike M (2005a) Auditory evoked magnetic fields in relation to iterated rippled noise. Hear Res. Jul;205:256-61.

Soeta Y, Nakagawa S, Tonoike M (2005b). Auditory evoked fields to variations of interaural time delay. Neurosci Lett. 383:311-6.

Soeta Y, Hotehama T, Nakagawa S, Tonoike M, Ando Y (2004) Auditory evoked magnetic fields in relation to interaural cross-correlation of band-pass noise. Hear Res. 196:109-14.

Sperling AJ, Lu ZL, Manis FR, Seidenberg MS. (2005) Deficits in perceptual noise exclusion in developmental dyslexia. Nature Neuroscience. 8:862-3.

Stein J (2001) The Magnocellular theory of developmental dyslexia. Dyslexia 7:12-36.

Stein J, Walsh V (1997) To see but not to read; the magnocellular theory of dyslexia. Trends in Neurosciences. 20:147-52.

Stern RM, Trahiotis C (1995) Models of binaural interaction. In: Hearing: handbook of perception and cognition, 2nd ed (Moore BCJ, ed.), pp. 347-386. San Diego: Academic Press

Stufflebeam SM, Poeppel D, Rowley HA, Roberts TP (1998) Peri-threshold encoding of stimulus frequency and intensity in the M100 latency Neuroreport 9:91-94.

Super H, Spekreijse H, Lamme VA (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat Neurosci. 4:304-10.

Tallal P, Miller S, Fitch RH. (1993) Neurobiological basis of speech: a case for the preeminence of temporal processing. Annals of the New York Academy of Sciences. 682:27-47.

Tallal P (1980). Auditory temporal perception, phonics and reading disabilities in children. Brain and Language. 9:182-198.

Tcheriev S (1904) Propriétés électromotrices du cerveau et du Coeur. J Physiol. Path. Gen 6 :671-682.

Tcheriev S(1902) Le téléphone comme indicateur d'une excitation nerveuse. J Physiol. Path. Gen 5 :671-682.

Temple E (2002) Brain mechanisms in normal and dyslexic readers. Current Opinion in Neurobiology. 12:178-83.

Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM, Gabrieli JD. (2000) Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. Proceedings of the National Academy of Sciences of the U S A. 97:13907-12.

Thompson KG, Hanes DP, Bichot NP, Schall JD (1996) Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. J Neurophysiol. 76:4040-55.

Tiitinen H, Mäkelä AM, Makinen V, May PJ, Alku P (2005) Disentangling the effects of phonation and articulation: hemispheric asymmetries in the auditory N1m response of the human brain. BMC Neurosci. 6:62.

Tillmann B, McAdams S (2004) Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. J Exp Psychol Learn Mem Cogn. 30:1131-42.

Torgesen JK, Wagner RK, Rashotte CA. (1994). Longitudinal studies of phonological Processing and reading. J. Learn. Disabil. 27:276-86; discussion 287-91.

Triesman A (1985) Preattentive processing in vision. Computer Vision, Graphics and Image Processing 31:156-177

Turkeltaub PE, Gareau L, Flowers DL, Zeffiro TA, Eden GF (2003) Development of eural mechanisms for reading. Nature Neuroscience 6:767-773.

Ulanovsky N, Las L, Farkas D, Nelken I. (2004) Multiple time scales of adaptation in auditory cortex neurons. J Neurosci. 24:10440-53.

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci. 6:391-8.

van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural
processing of auditory speech. Proc Natl Acad Sci U S A. 102:1181-6.

Vecera SP, Flevaris AV, Filapek JC (2004) Exogenous spatial attention influences
figure-ground assignement. Psychol Sci. 15:20-6.

Wang X, Lu T, Liang L (2003) Cortical Processing of Temporal Modulations.
Speech Communication 41:107-121.

Wechsler D(1999). Wechsler Abbreviated Scale of Intelligence. San Antonio, TX:
Psychological Corporation.

Wiederholt JL, Bryant BR (1992). Gray Oral Reading Tests, Third Edition.
 Austin, TX: Pro-Ed.

Winkler I (2003) Change detection in complex auditory environment: beyond the
oddball paradigm. In Polich, J (ed): Detection of change: event related
potential and fMRI findings. Boston, USA: Kluwer academic publishers. Pp:
1:22

Winkler I, Karmos G, Naatanen R (1996) Adaptive modeling of the unattended
acoustic environment reflected in the mismatch negativity event-related
potential. Brain Res. 742:239-52.

Witton C, Talcott JB, Hansen PC, Richardson AJ, Griffiths TD, Rees A, Stein JF,
Green GG (1998) Sensitivity to dynamic auditory and visual stimuli predicts
nonword reading ability in both dyslexic and normal readers. Current Biology
8:791-7.

Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE
(1993) Modulation of early sensory processing in human auditory cortex

during auditory selective attention. Proc Natl Acad Sci U S A. 90:8722-6.

Wolfe JM, Kluender KR, Levi DM, Bartoshuk LM, Herz RS, Klatzky RL, Lederman SJ (2006) Sensation and perception. Sunderland, MA: Sinauer publishers.

Wood FB, Grigorenko EL (2001). Emerging issues in the genetics of dyslexia: A methodological preview. Journal of Learning Disabilities, 34(6): 503-511.

Woodcock RW, McGrew KS, Mather N. (2001). Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside Publishing.

Woodcock RW, Johnson M.G. (1977). Woodcock Johnson Psycho-Educational Battery. Allen, TX: DLM.

Woods DL, Alain C, Covarrubias D, Zaidel O (1995) Middle latency auditory evoked potentials to tones of different frequency. Hear Res. 85:69-75.

Wright BA, Bowen RW, Zecker SG. (2000) Nonlinguistic perceptual deficits associated with reading and language disorders. Current Opinion in Neurobiology. 10:482-6.

Wright BA, Lombardino LJ, King WM, Puranik CS, Leonard CM, Merzenich MM. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. Nature. 387:176-8.

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

Yost WA, Popper AN, Fay RR (1993) Human psychophysics. NY, NY: Springer-Verlag

Yost WA, Harder PJ, Dye RH (1987) Complex spectral patterns with interaural

differences: dichotic pitch and the 'central spectrum". In: Auditory processing of complex sounds (Yost WA et al., eds.), pp. 190-201.NJ: Erlbaum, Hillsdale.

Yvert B, Crouzeix A, Bertrand O, Seither-Preisler A, Pantev C (2001) Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. Cereb. Cortex 11:411-423.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science. 256:846-9.

Zatorre RJ (2001) Neural specializations for tonal processing. Ann N Y Acad Sci 930: 193-210.

Zhang PX, Hartmann WM (2004) On the lateralization of Huggins pitches. J. Acoust. Soc. Am. 115:2534.

Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C. (2005) Deficits in speech perception predict language learning impairment. Proceeding of the National Academy of Sciences of the U S A. 102:14110-5.

Zipser K, Lamme VA, Schiller PH (1996) Contextual modulation in primary visual cortex.. J Neurosci. 16:7376-89.