The processing of temporal fine-structure information in the human auditory system

by David Amooti Magezi, MA(Cantab), BM BCh (Oxon)

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy, March 2010

Acknowledgements

First, I thank Dr Katrin Krumbholz,

For excellent support, dedication and brilliant ideas,

Always available for questions and

Enthusiastic discussion, over the 3+ years,

For teaching the art

Of communicating science to peers,

And for making it such a pleasure to research

The system between "between our two ears".

For funding support, I thank

The Medical Research Council, MRC.

For technical assistance with

fMRI data collection (see Chapter 3),

I thank Kay Head.

For help with psychoacoustic data collection

I thank Rosanna Moore

And Sara Ponting (see Chapter 4),

As well as Imogen BoSmith (see Chapter 5).

I also thank all in the MRC Institute of Hearing Research

For making it an environment in which to thrive.

Nick Clark, Barrie Edmonds and Paul Briley,

I thank for the friendly office atmosphere

And for many a Matlab tip.

And for all the data that you see here,

I must thank all the participants

Who for these studies, did volunteer.

For registration and general research training, I thank the School of Psychology

And Dr Martin Schürman.

For unconditional support, I thank my mother, "Webaale Amooti",

For encouragement, patience and proof-reading, I thank Akiiki.

Abstract

The auditory nerve conveys fine-grained temporal information that reflects individual cycles of the basilar membrane vibration. The current project is concerned with how this temporal fine-structure information is processed in the human auditory system. Integration of fine-structure temporal information across the ears (binaural processing) plays a crucial role in sound localisation and signal detection in noise. However, in monaural processing, the role of temporal fine-structure information remains uncertain, because spectral information is usually also available.

The first study in this project used behavioural methods, along with model simulations, to show that the binaural system exploits phase differences between disparate frequency channels for processing fine-structure interaural temporal differences (ITDs). The second study explored the neural representation of ITDs by using electroencephalography (EEG) to measure the transient brain response to a change in ITD in an otherwise continuous sound. The results suggest that fine-structure ITDs are coded by a non-topographic opponent-channel mechanism, based on the overall activity levels in two broadly tuned hemispheric channels. The third study used rapid event-related functional magnetic resonance imaging (fMRI) to investigate the topography of the transient ITD change response measured in the second study. The ITD change response was compared with the transient response to the onset of pitch in an otherwise continuous sound. It was found that the topographies of the transient ITD and pitch responses were very similar to the topographies of the corresponding sustained responses measured in previous epoch-related fMRI studies.

The last two studies examined whether temporal fine-structure information is used for frequency coding in monaural processing. The fourth study aimed to eliminate temporal fine-structure cues from the neural representation of low-frequency pure tones by presenting the tones in conditions of binaural unmasking, because a previous study had shown that temporal envelope cues to pitch are inaccessible in such masking conditions. However, frequency discrimination performance for pure tones was found to be similar in monaural and binaural masking conditions. The fifth study suggests that this was because frequency discrimination of low-frequency pure tones relies on spectral rather than temporal cues. In this study, frequency discrimination performance was measured for partially masked pure tones and was found to reflect the level-dependent changes in the shape of the pure-tone excitation pattern.

Table of contents

List of abbreviations

1I2AFC	one-interval two-alternate forced-choice
2I2AFC	two-interval two-alternate forced-choice
A	amplitude
A_{crit}	criterion area
AFz	central forehead electrode
Ag/AgCl	silver/silver chloride
AM	amplitude modulation
ANOVA	analysis of variance
APD	axonal propagation delay
BESA	brain electrical source analysis software
BMLD	binaural masking level difference
BOLD	blood oxygen level-dependent
CB	central band
CF	characteristic frequency
cN1, cP1, cP2	deflections of the ITD change response
CR	change response
CSP	continuous stimulation paradigm
Cz	vertex electrode
D	decision measure
D_{crit}	criterion value of D
dB	decibel
DC	direct current

discr.	discrimination
DSAM/AMS	auditory modelling software package (O'Mard and Meddis, 2004)
ERB	equivalent rectangular bandwidth (Glasberg and Moore, 1990)
EEG	electroencephalography
EEGLAB	EEG analysis software (Delorme and Makeig, 2004)
EOR	energy onset response
EPI	echo-planar imaging
F_c	cutoff frequency of filter
F_{car}	carrier frequency
F_{mod}	modulation rate
FFT	fast Fourier transform
FL	Florida
FM	frequency modulation
fMRI	functional magnetic resonance imaging
F_s	signal frequency
FSR	failure-to-success ratio
HG	Heschl's gyrus
HS	Heschl's sulcus
Hz	hertz
IC	inferior colliculus
ICBM	international consortium for brain mapping
IIR	infinite impulse response
IRN	iterated rippled noise

ITD	interaural time difference
kHz	kilohertz
LI	laterality index
m	modulation index
MA	Massachusetts
MEG	magnetoencephalography
ML	motion-left condition
MR	motion-right condition
mm	millimetres
MM	mixed modulation
MNI	Montreal Neurological Institute
MPRAGE	magnetization prepared rapid gradient echo
MRC	Medical Research Council
ms	milliseconds
MSO	medial superior olive
N	null condition
N_0S_0	homophasic masking condition
N_0S_π	antiphasic masking condition
nAm	nanoamperes
NAP	neural activity pattern
OffR	offset response
P	pitch condition
PLSD	probable least significant difference

pp	peak-to-peak
PP	planum polare
PT	planum temporale
rms	root mean square
s	signal
SCB	slip cycle band
SL	sensation level
SOA	stimulus onset asynchrony
SPL	sound pressure level
SPM	"statistical parametric mapping" software package (http://www.fil.ion.ucl.ac.uk/spm).
SR	sustained response
STP	supratemporal plane
t	time
T	Tesla
TDT	Tucker-Davies Technologies
TE	echo time
TE1.0-TE1.2	cyto-architechtonic subdivisions of primary auditory cortex
thr.	threshold
TI	inversion time
TR	image repetition time
XCH	cross-channel
α	compression exponent

ΔF	frequency difference
ΔF_c	frequency excursion (maximum- minimum frequency)
ΔITD	ITD difference
µs	microseconds
µV	microvolts
ϕ_{car}	starting phase of carrier
ϕ_{AM}	starting phase of AM
ϕ_{FM}	starting phase of FM

General introduction

The auditory system can process stimulus-related temporal information with an acuity that is unrivalled in the mammalian brain (for review, see Oertel, 1997, 1999). The basilar membrane vibrates in the rhythm of the temporal waveform of the stimulating sound, and this leads to action potentials being generated in the auditory nerve fibres. If the sound frequency is not too high, the action potentials are time-locked to the individual basilar membrane deflections, and the resulting timing information is referred to as temporal fine-structure information. In humans, the comparison of temporal fine-structure information across the two ears plays a crucial role in low-frequency sound localisation and helps to perceive sounds in noisy environments (Licklider, 1948; Wightman and Kistler, 1992; Lavandier and Culling, 2008; for review, see Durlach and Colburn, 1978). Temporal fine-structure information is also assumed to be processed monaurally, for example via a process of autocorrelation, to encode sound frequency and complex pitch (for review, see Moore, 2008).

The aim of the current project was to use psychoacoustics, electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) to investigate how the human auditory system processes temporal fine-structure information for the perception of pitch and sound lateralisation mediated by interaural time differences (ITDs).

Most computational models of binaural perception (e.g. Stern and Trahiotis, 1995) are based on the traditional delay-and-coincidence theory of interaural temporal processing (Jeffress, 1948). An important assumption of these traditional models is that the binaural system only compares temporal fine-structure information from

corresponding frequency channels. This is because, due to the travelling-wave nature of the cochlear phase response, a mismatch between channels would add an internal phase delay to the external ITD and would thus distort the ITD estimate. Furthermore, these models also assume an array of neurons tuned to ITDs within the physiological range, which is the range of ITDs experienced in natural environments, as determined by head size. However, physiological results have shown that, in small-headed mammals, a significant proportion of ITD-sensitive neurons are tuned to ITDs that lie outside of the physiological range (McAlpine et al., 1996, 2001; McAlpine, 2005). Moreover, the same results also indicate that the distribution of the best ITDs of binaural neurons is highly dependent on the neurons' best frequencies, whereas delay-and-coincidence models would predict ITD tuning to be independent of frequency (see McAlpine et al., 2001). Alternative models have been suggested to explain these data. One model suggests that it might be the cochlear phase delays between disparate frequency channels (Schroeder, 1977; Shamma et al., 1989; Joris 2004, 2006), rather than axonal propagation delays between corresponding channels, that create the internal delays for binaural coincidence neurons. This idea was tested in the first study of the current project, which is described in **Chapter 1**. In this study, ITD discrimination thresholds were measured for a pure tone that was partially masked by a highpass-filtered noise in one ear, and a lowpass-filtered noise in the other ear. Cross-channel models of ITD processing would predict that listeners would be able to extract ITDs from such asymmetrically masked tones, whereas models that only assume comparisons between corresponding channels, such as delay-and-coincidence models, would predict the task to be difficult, if not impossible.

Another suggestion is that ITDs are coded by the overall activity levels in two opponent neural populations, or channels, which are broadly tuned to the two acoustic hemifields (von Békésy, 1930; van Bergeijk, 1962; McAlpine et al., 2001; Brand et al., 2002; Stecker et al., 2005), rather than by the spatial, or topographic, distribution of activity across many finely tuned channels, as assumed in delay-and-coincidence models. This suggestion was investigated in the second study, described in **Chapter 2**, which measured the electroencephalographic (EEG) response to an abrupt change in ITD in an otherwise continuous noise stimulus. The ITD change was either away from ("outward" change) or towards the midline ("inward" change). According to the opponent-channel model, the response to an outward ITD change should be much larger than the response to the corresponding inward change, whereas topographic models would predict similar response sizes in both conditions.

The response to an ITD change in an otherwise continuous sound, as measured in the second study, would be assumed to reflect the response of those neural elements that are specifically involved in the processing of interaural temporal information. However, the response could also be due to an unspecific change detection mechanism of the kind suggested to underlie the auditory oddball or mismatch response (for review, see Näätänen and Winkler, 1999), or to a generic "edge" detection process, related to the perception of auditory objects, as suggested by Chait et al. (2008; see also Chait et al., 2007). In order to investigate this question, the third study in this project, described in **Chapter 3,** used fMRI to measure the topography of the transient ITD change response used in the second study. The ITD change response was acquired with a rapid event-related design and a meta-analysis was performed to compare its topography with that of

ITD-specific fMRI responses obtained with conventional epoch-related designs. For comparison, the study also included a pitch condition, because pitch and ITD processing have been shown to activate different areas in epoch-related studies (Warren and Griffiths, 2003; Barrett and Hall, 2006; for review, see Arnott et al., 2004).

While temporal fine-structure information is known to play a crucial role in binaural hearing, its role in monaural processing remains uncertain, because sounds that convey temporal fine-structure information also convey spectral information. Spectral information is mediated by the distribution of activation strength across the tonotopic array, referred to as the excitation pattern (Zwicker, 1956, 1970). The last two studies in the current project, described in **Chapters 4 and 5**, aimed to address the question of whether frequency is coded by temporal fine-structure information or by spectral information. The fourth study (**Chapter 4**) was inspired by the finding that binaural sluggishness eliminates temporal envelope cues to pitch in conditions of binaural unmasking (Krumbholz et al., 2009a). In harmonic tones, temporal envelope cues arise as a result of harmonic beating within individual cochlear filters. The aim of the current study was to investigate whether binaural sluggishness would also affect the faster-varying temporal fine-structure information. For that, frequency discrimination performance for pure tones was measured in conditions of binaural unmasking. The hypothesis was that, if binaural sluggishness degrades fine-structure temporal cues and if frequency coding is based on these fine-structure cues, frequency discrimination performance would be expected to be severely impaired in conditions of binaural unmasking. In contrast, if frequency were coded spectrally, frequency discrimination

performance in conditions of binaural unmasking would be expected to be similar to performance in diotic masking conditions.

The fifth study (**Chapter 5**) measured frequency discrimination performance in partially masked pure tones. Highpass and lowpass noise maskers were used to obscure either the low- or high-frequency flank of the tones' excitation pattern, and the slopes of these excitation pattern flanks were then manipulated by varying sound level (Egan and Hake, 1950; Ruggero et al., 1997; for review, see Robles and Ruggero, 2001). If frequency were coded by a spectral mechanism, performance would be expected to reflect the level-dependent changes in the slopes of the excitation-pattern flanks. In contrast, if frequency were coded temporally, performance should be independent of the shape of the excitation pattern (Moore and Sek, 1996).

Chapter 1. Can the binaural system extract fine-structure interaural time differences from non-corresponding frequency channels?[1]

1.A INTRODUCTION

In humans, horizontal sound localization is mainly based on the microsecond differences in sound arrival time between the two ears (interaural time differences, ITDs), produced by the path-length differences between the ears and the sound source. ITDs are processed by a comparison of the phase-locked temporal information mediated by the left and right auditory nerves. As the basilar membrane response to sound is a travelling wave, the phase of the response changes as a function of place along the membrane. This is why most models of binaural processing make the assumption that ITDs are processed in a channel-by-channel manner and that the channels, the temporal responses of which are being compared, originate from corresponding places in the two cochleae (Colburn, 1996). Jeffress (1948), for instance, proposed that ITDs are processed by means of a delay-and-coincidence mechanism, in which the signals from corresponding left- and right-ear channels are delayed relative to each other by axonal delays and then converge onto neurons that are excited only by coincident input.

As the slope of the cochlear phase change is particularly steep around the point where the travelling wave reaches its maximum (i.e., where the auditory response is most sensitive), the spatial correspondence between converging left- and right-ear channels would need to be very precise. A mismatch between channels would add an internal phase difference to the external ITD and would thus distort the ITD estimate. Bonham and Lewis (1999) used a gammatone-filter model based on auditory-nerve data in the cat to show that the allowable degree of mismatch between channels is 0.012 octaves (about

[1] Based on Magezi and Krumbholz (2008)

4.2 Hz) at 500 Hz. According to Greenwood's (1990) cochlear frequency-position function, this corresponds to a cochlear distance of only about 3 hair cells, implying quite a remarkable degree of precision required in the synaptic innervation of binaural neurons. During development, this kind of precision would be assumed to be achieved by a process of activity-dependent pruning of an initial, larger set of connections containing both matched and unmatched pairs (for review, see Friauf and Lohmann, 1999). The pruning would be assumed to be based on the response phases of the channel pairs. In the gammatone filter, the most widely used auditory-filter model (Patterson, 1994), response phase changes by more than 180° within one equivalent rectangular filter bandwidth (ERB; Glasberg and Moore, 1990). Any local mismatch in channel alignment would require a recalibration of the internal delay mechanism (e.g., axonal delay lines) at the relevant frequency.

While cochlear phase delays constitute a problem for models like the Jeffress model, it has been suggested that these phase delays might actually have a positive role in binaural temporal processing. Schroeder (1977) proposed that it might be the phase delays between non-corresponding cochlear channels, rather than axonal delays between corresponding channels, that create the internal delays for binaural coincidence neurons. This idea was tested and expanded by Shamma et al. (1989) in a computational implementation, referred to as the "stereausis" model. The stereausis model assumes that there are binaural connections between both corresponding and non-corresponding channels. Thus, in this case, the developmental pruning process could be much coarser than in the case of the Jeffress or related models (e.g., driven by between-channel differences in overall activity level, rather than response phase). In this case, the fine-

tuning of the mechanism would be assumed to occur at a higher stage, where the binaural activity patterns are interpreted, rather than at the stage where the patterns are generated, and would not require calibration with neurally generated delays.

An argument against the stereausis mechanism is that it may be susceptible to the changes in the cochlear phase response that are known to result from changes in sound level (Anderson et al., 1971; Carney and Yin, 1988; Nuttall and Dolan, 1993; Ruggero et al., 1997); and that would also be expected to occur as a result of cochlear damage. Moreover, the stereausis model would not be readily applicable to ITDs in the temporal envelope of high-frequency sounds, because the temporal envelope does not exhibit the same phase differences across channels as the temporal fine structure does (e.g., Carlyon and Shamma, 2003). Thus, envelope ITDs would have to be assumed to be processed by a different mechanism than fine-structure ITDs, and this assumption is consistent with physiological results suggesting that interaural cues in low- and high-frequency sounds are processed in different structures (Joris and Yin, 1995; Batra et al., 1997; for review, see Tollin 2003; see however Griffin et al., 2005). Furthermore, although psychophysical studies show that envelope ITDs are processed with a similar accuracy as fine-structure ITDs (van der Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002, 2003), this is only true for relatively low stimulus frequencies; above 300 Hz accuracy rapidly deteriorates for envelope but not fine-structure ITDs (Bernstein and Trahiotis, 2002).

In favour of the stereausis model, Joris et al. (2004, 2006) pointed out that the presence of internal delays outside the physiological range (i.e., the range of ITDs encountered for a given head size; McAlpine et al., 1996, 2001; Fitzpatrick et al., 2000; Brand et al., 2002) could be better explained by a stereausis-type model than by a

Jeffress-type model. Joris et al. (2004,2006) also showed that, in the cat, the dependence of the best ITD of binaural neurons in the inferior colliculus (IC) on the neurons' characteristic frequency (CF; McAlpine et al., 1996) is similar to the dependence on CF of the phase differences between responses from non-corresponding auditory nerve fibres. However, Brand et al. (2002) proposed that internal delays beyond the physiological range could also be created by neural mechanisms, rather than cochlear disparities. Their results suggest that internal delays are generated by contralateral inhibitory input to binaural neurons in the medial superior olive (MSO), which slightly precedes the excitatory input from the same side. Whether inhibition could explain the sometimes very large best ITDs of low-CF binaural neurons is currently a matter of debate (Joris and Yin, 2007).

A further possibility is that cross-channel comparisons may be part of a mechanism based on neurally generated internal delays, such as the Jeffress mechanism, to make it more robust against channel mismatches. For instance, in a Jeffress-type mechanism, the overall ITD could be derived by an activity-weighted average of estimates from both matched and unmatched channel comparisons. It could also be the case that both neurally generated and cochlear phase delays contribute to the internal delays for ITD processing. This idea is supported by computational analyses, which suggest that both kinds of delay may be needed to explain the ITD sensitivity of binaural neurons in mammals (Bonham and Lewis, 1999; Zhou et al., 2005).

If internal delays for ITD processing are produced by stereausis, the auditory system would be expected to be able to extract ITDs from disparate channels, which can be tested by making information from corresponding channels unavailable. This approach

has been used in several psychoacoustical studies, with varied results. Some studies suggest that ITDs can only be extracted from corresponding channels (Toole and Sayers, 1965b); other studies seem to show that ITDs can also be extracted from non-corresponding channels (Schubert and Elpern, 1959; Zerlin, 1969), but only when the frequency separation of the channels is less than one critical band (Scharf et al., 1976), and yet other studies suggest that ITDs can be extracted from channels even with fairly large frequency separations (Deatherage, 1961, 1966). The main problem with all of these studies is that they used ITDs in the temporal envelopes of higher-frequency sounds, to which a stereausis-type mechanism would not be readily applicable.

The aim of the current study was to address the question of whether the auditory system is able to extract fine-structure ITDs from non-corresponding cochlear channels. For that, we measured the ITD discrimination threshold for 500-Hz pure tones, which were partially masked by a lowpass-filtered noise in one ear and a highpass-filtered noise in the other ear. The assumption was that the highpass noise would mask the basal part of the cochlear response to the tone, while the lowpass noise would mask the apical part of the response, thus forcing the listener to extract ITDs from disparate channels. The ITD discrimination threshold for these "dichotically-masked" tones was compared to the threshold for "diotically-masked" tones, i.e., tones that were masked by the same type of noise (low- or highpass) in both ears. We expected models based on a comparison of corresponding channels, such as the Jeffress model, to predict that ITD discrimination for the dichotically-masked tones would be more difficult than for the diotically-masked tones, which the data showed to indeed be the case; in contrast, models based on cross-channel comparisons, such as the stereausis model, were expected to predict ITD

discrimination performance for the dichotic masking conditions to be similar to that for the diotic conditions. While initial modelling confirmed this expectation, and thus favoured the Jeffress over the stereausis model, further simulations revealed that a more physiologically-plausible version of the stereausis model was also able to predict the observed threshold difference between the diotic and the dichotic conditions. Moreover, this modified stereausis model was able to account for individual aspects of the data, which the Jeffress model was unable to predict.

1.B METHODS

Stimuli

ITD discrimination thresholds were measured for 500-Hz pure tones, partially masked by two independent (uncorrelated), continuous noise maskers, each presented to one ear. The experiment comprised three 'diotic' masking conditions, in which the spectral composition of the maskers was the same in the two ears, and two 'dichotic' conditions, in which the spectral composition of the maskers differed between the ears. The diotic conditions comprised a "lowpass", a "highpass" and an "allpass" condition. In the lowpass and highpass conditions, the maskers in both ears were low- or highpass filtered, respectively, with the same cutoff frequency. In the "allpass" condition, both maskers were neither low- nor highpass filtered. In the dichotic conditions, the masker was lowpass filtered in one ear and highpass filtered in the other ear; the lowpass masker was presented to either the left ('dichotic left') or the right ear ('dichotic right'). All maskers were filtered to produce a roughly constant level of excitation within their passbands. The stimuli were generated digitally at a sampling rate of 25 kHz using TDT System 3 (Tucker-Davies Technology, Alachua, FL, USA) and MATLAB® (The

Mathworks, Natick, MA, USA). Filtering of the noise maskers was carried out in the spectral domain using 2^{18}-point fast Fourier transforms (FFTs). The low- and highpass filters were implemented as brick wall filters (i.e., setting all spectral components outside the passband to zero).

In order to maximize the chances of detecting any effect of masking condition, the ITD discrimination threshold was measured for a range of four different sensation levels (6, 8, 12 and 16 dB SL), which were the same for all masking conditions. Any effect of masking condition would be expected to disappear towards higher sensation levels, because all parts of the excitation pattern of the tone would eventually become audible at higher levels, irrespective of the spectral composition of the masker. Conversely, at very low sensation levels, any effect of masking condition might be masked by floor effects (i.e., the general difficulty of performing the task at low levels). The sensation level of the tone could be varied either by varying the level of the tone or the masker or, in the case of the low- and highpass maskers, by changing the cutoff frequency of the filter. As level changes can alter the transfer characteristics of the cochlear filters, sensation level was varied by adjusting the masker cutoff frequency in the case of the low- and highpass maskers, and by changing the masker level in the case of the allpass masker. For that, the detection threshold of the tone was first determined for all diotic masking conditions (lowpass, highpass and allpass) and in quiet. In the low- and highpass conditions, detection threshold was measured as a function of filter cutoff frequency. Cutoff frequencies were smaller than or equal to 500 Hz in the lowpass condition, and larger than or equal to 500 Hz in the highpass condition. In order to sample the relevant parts of the masking patterns, four different cutoff frequencies were used in both the low- and

highpass conditions, with distances of 0, 0.5, 1 and 1.5 ERBs from 500 Hz (Fig. 1.1; the corresponding frequency differences in hertz are shown on the top axis). In the detection threshold measurements, all maskers had the same level of about 55 dB SPL per ERB within their passbands.

For the ITD discrimination threshold measurements, the sound pressure level of the tone was set to 6 dB above the detection threshold for the lowpass masker with the 500-Hz cutoff (0 ERBs), yielding an average tone level of 47.4 dB SPL (std. dev. = 2.3 dB). The cutoff frequencies of the lowpass and highpass maskers were then set to yield the desired tone sensation levels (6, 8, 12 and 16 dB SL) by linearly interpolating the detection threshold function for the respective masker type (Fig. 1.1). The resulting cutoff frequencies were used in the diotic low- and highpass conditions and in the dichotic conditions. The level of the low- and highpass maskers was the same as in the detection threshold measurements (55 dB per ERB). In the allpass condition, the desired sensation levels were achieved by reducing the masker level by the appropriate amount (based on the presentation level of the tone and its detection threshold in the 55-dB SPL allpass masker). Tone duration was 500 ms including 10-ms squared-cosine on- and off-ramps.

Stimuli were digital-to-analogue converted with a 24-bit resolution (TDT RP2.1), amplified (TDT HB7) and presented over headphones (K240 DF, AKG, Vienna, Austria) to the participant, who was seated in a double walled sound-attenuating room.

Procedure

Both the detection and ITD discrimination threshold measurements used an adaptive two-interval, two-alternative forced-choice (2I2AFC) procedure with a three-down, one-up rule, which tracks 79% correct performance (Levitt, 1971). The two

observation intervals in each trial had a duration of 500 ms and were separated by a gap of 500 ms. Visual feedback was provided at the end of each trial.

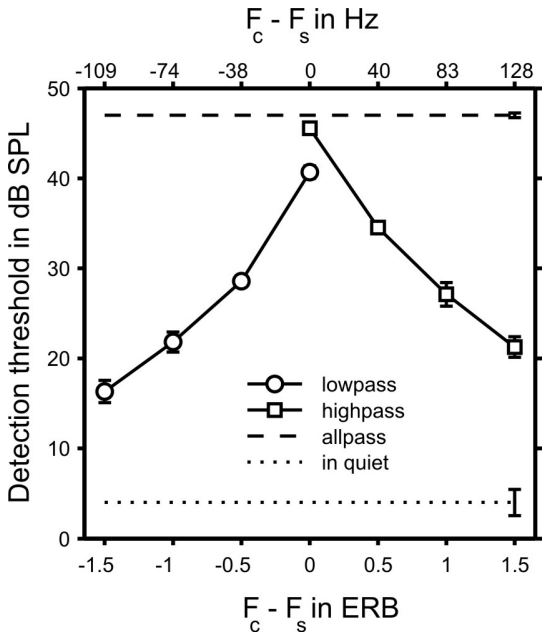

FIG. 1.1 Masked detection threshold for a 500-Hz pure tone plotted as a function of the frequency separation between the masker cutoff frequency, F_c, and the signal frequency, F_s, of 500 Hz in ERBs (bottom axis) or hertz (top axis; positive and negative values represent masker cutoff frequencies above and below 500 Hz, respectively). The circles and squares show the thresholds for the lowpass and highpass masking conditions, respectively. The dashed horizontal line shows the threshold for the allpass masking condition, and the dotted line shows the threshold in quiet. Thresholds were averaged across five participants; error bars show the standard error of the mean.

In the detection threshold measurements, only one of the two intervals contained a tone; any masking noise was presented continuously throughout the entire threshold run. The task was to indicate which of the two intervals contained the tone by pressing one of two response buttons. The level of the tone was changed in steps of 5 dB up to the first reversal in level, 3 dB up to the second reversal and 2 dB for the rest of the 12 reversals that made up each threshold run. Each threshold estimate was taken as the arithmetic mean of the levels at the last 10 reversals. At least three such threshold estimates were averaged to obtain the final threshold for each condition. The order in which different conditions were tested was randomized for each of the three threshold runs.

In the ITD discrimination threshold measurements, both intervals contained a tone. The tones were gated on and off synchronously at the two ears, but their fine-structure contained an ITD. The ITD was of the same magnitude in both intervals, and was leading at the right ear in one interval and at the left ear in the other interval. The task was to indicate which of the two intervals contained the rightmost sound. The ITD difference between the two tones, ΔITD, was reduced and increased by a factor, v. Up to the first reversal, v was equal to 2; it was reduced to 1.5 up to the second reversal and set to 1.3 for the following 10 reversals. Each threshold estimate was taken as the geometric mean of the ITD difference, ΔITD, for the last 10 reversals. The ITD discrimination threshold for each condition is the mean of at least three such threshold estimates. As for the detection thresholds, the order in which conditions were tested was randomized. If, during the adaptive track, ΔITD exceeded 200 μs (one tenth of the period of the 500-Hz

signal) on three consecutive trials, the run was terminated and no threshold estimate was recorded for that run.

Data analysis

The individual detection thresholds were submitted to a two-way repeated-measures ANOVA with masking condition (low- and highpass) and distance of masker cutoff frequency from the 500-Hz signal frequency (0, 0.5, 1 and 1.5 ERBs) as independent within-participant factors. To test for individual- and group-level effects in the ITD discrimination threshold data, individual and average threshold estimates of individual participants were submitted to two-way repeated-measures ANOVAs with masking condition (lowpass, highpass, allpass, dichotic left and right) and tone sensation level (6, 8, 12 and 16 dB SL) as independent within-participant factors.

Participants

A total of seven participants were initially recruited, five of whom (1 male and 4 female, aged between 24 and 44 years) completed the study. All participants had absolute thresholds of 25 dB HL or less at audiometric frequencies, and had no history of hearing or neurological disorders. One of the two participants who did not complete the study became unavailable, and the other remained unable to reliably attain a threshold estimate in one of the most difficult conditions (dichotic left at 6 dB SL) even after extensive training. Participants who were not authors of the corresponding manuscript (Magezi and Krumbholz, 2008) were paid for their services at an hourly rate. Two of the five participants who completed the study were experienced in ITD discrimination and could do the task without any prior practice. The other participants underwent between 4 and 15 hours of training, starting with the easiest conditions (the diotic conditions at 16 dB SL)

and gradually moving on to the more difficult ones. The experimental procedures were approved by the Ethics Committee of the Nottingham University School of Psychology.

Additional ITD matching experiment

During the ITD discrimination task with the dichotic maskers, participants reported perceiving the tone as being lateralized towards the ear that received the lowpass masker (see Sec. 1.C). In order to verify these subjective reports more formally and quantify the degree of lateralization produced by the dichotic maskers, an additional experiment was conducted to measure the external ITD necessary to bring the dichotically masked tones to midline. This "matching ITD" was determined with an adaptive one-interval, two-alternative forced-choice (1I2AFC) procedure. On each trial, a single 500-ms tone was presented and the task was to indicate whether the tone was perceived to the left or the right of the midline. The tone contained an ITD, which was changed adaptively according to the participants' responses. As this was a subjective task, no feedback was provided.

Each matching run consisted of two adaptive tracks, which were randomly interleaved. One of the tracks, referred to as the "down track", used a two-down one-up rule (the ITD of the tone was decreased after two consecutive "right" responses and increased after each "left" response) to estimate the ITD that would yield 70.7% "right" responses. The other track, referred to as the "up track", used a two-up one-down rule (the ITD was increased after two consecutive "left" responses and decreased after each "right" response) to estimate the ITD yielding 70.7% "left" responses. The starting ITD was +500 μs for the down track and −500 μs for the up track. In both tracks, the step size of the ITD increases and decreases was 250 μs up to the first reversal in ITD, 125 μs up

to the second reversal and 62.5 μs for the rest of the 8 reversals that made up each track. The ITDs at the last 6 reversals of both tracks were averaged to yield an estimate of the matching ITD. At least three such estimates were averaged to obtain the matching ITD for each condition. The order in which different conditions were tested was randomized. If, during either of the adaptive tracks, the magnitude of the ITD exceeded 1000 μs on three consecutive trials, the run was terminated and no matching ITD estimate was recorded for that run.

At the time when the ITD matching experiment was conducted, only three of the five participants who had taken part in the main experiment were still available for testing. For these three participants, the masking conditions tested in the ITD matching experiment were identical to the dichotic masking conditions used in the ITD discrimination threshold measurements of the main experiment (lowpass masker presented either to the left or right ear; masker cutoff frequencies set to yield four different tone sensation levels of 6, 8, 12 and 16 dB). In order to verify the results of these three participants, five new participants (3 male, 2 female, aged between 23 and 34 years, with normal hearing at audiometric frequencies), four of whom were experienced in psychoacoustical tasks, were recruited to this experiment. For these new participants, the matching ITD was measured for only one tone sensation level of approximately 10 dB. In this case, the masker cutoff frequencies were not based on individual detection threshold data, but were derived from the average detection threshold function measured in the main experiment. The procedure to derive the cutoff frequencies was the same as in the main experiment. The cutoff frequencies of the lowpass and highpass maskers were separated from the signal frequency of 500 Hz by 0 and 0.219 ERBs (0 and 17.4 Hz),

respectively, and the tone was presented at a level of 50.7 dB. Neither the three original nor the five new participants needed any training to perform the ITD matching task. General aspects of the stimulus presentation were the same as in the ITD discrimination measurements of the main experiment.

The matching ITDs of the three participants who had also taken part in the main experiment were submitted to a two-way repeated-measures ANOVA with masking condition (dichotic left and right) and tone sensation level (6, 8, 12 and 16 dB) as independent within-participant factors. The matching ITDs of the five new participants were submitted to a paired *t*-test.

1.C RESULTS AND INTERIM DISCUSSION

Group results

The detection thresholds were very consistent across participants, and so, only the average thresholds are shown in Fig. 1.1. In Fig. 1.1, the average thresholds for the lowpass and highpass conditions are plotted as a function of the difference between the masker cutoff frequency, F_c, and the signal frequency, F_s (500 Hz), in ERBs. When the masker cutoff frequency was equal to the signal frequency (F_c–F_s = 0 ERBs), the detection threshold for the highpass condition (squares in Fig. 1.1) was similar to that of the allpass condition (dashed horizontal line), and the threshold for the lowpass condition (circles) was on average 4.9 dB lower than for the highpass condition [main effect of masking condition: $F(1,4) = 15.236$, $p = 0.017$]. This asymmetry between the lowpass and highpass conditions may be related to the reported asymmetry in psychophysical suppression (Houtgast, 1972, 1973; Shannon, 1976), wherein higher frequencies are more effective at suppressing lower frequencies than vice versa. The asymmetry may also be

due to the presence of nonlinear distortion products at and around the signal frequency, which would be expected to be present in the highpass but not the lowpass condition (Plomp, 1965; Greenwood, 1971; Wiegrebe and Patterson, 1999). While the phenomenon of the upward spread of masking would predict a threshold asymmetry in the opposite direction, the masker level used in the current experiment would be deemed too low to elicit this effect (Egan and Hake, 1950). As expected, the detection threshold in the lowpass and highpass conditions decreased with increasing separation between the signal and the filter cutoff frequency [$F(3,12) = 2362.471$, $p < 0.001$]. There was no significant interaction between the main effects of masking condition and signal-to-masker-cutoff separation (F_c–F_s). The detection threshold for the allpass condition (dashed horizontal line) was an average of 43 dB higher than the detection threshold in quiet (dotted horizontal line).

The ITD discrimination threshold data revealed significant main effects of both sensation level [$F(3,12) = 129.887$, $p < 0.001$] and masking condition [$F(4,16) = 20.788$, $p < 0.001$] in the group-level analysis (see Sec. 1.B; Fig. 1.2). According to Fisher's probable least significant difference (PLSD) *post hoc* tests, the main effect of masking condition was mainly due to the thresholds for the dichotic conditions (dichotic left and right) being larger than those for the diotic conditions (lowpass, highpass and allpass; compare filled and open symbols in Fig. 1.2; $p < 0.001$, for all diotic-dichotic comparisons, except for the comparison between dichotic left and highpass, where $p = 0.006$). The main effect of sensation level was due to a substantial improvement in ITD discrimination performance with increasing sensation level, as would be expected. There

was no significant interaction between the main effects of masking condition and sensation level.

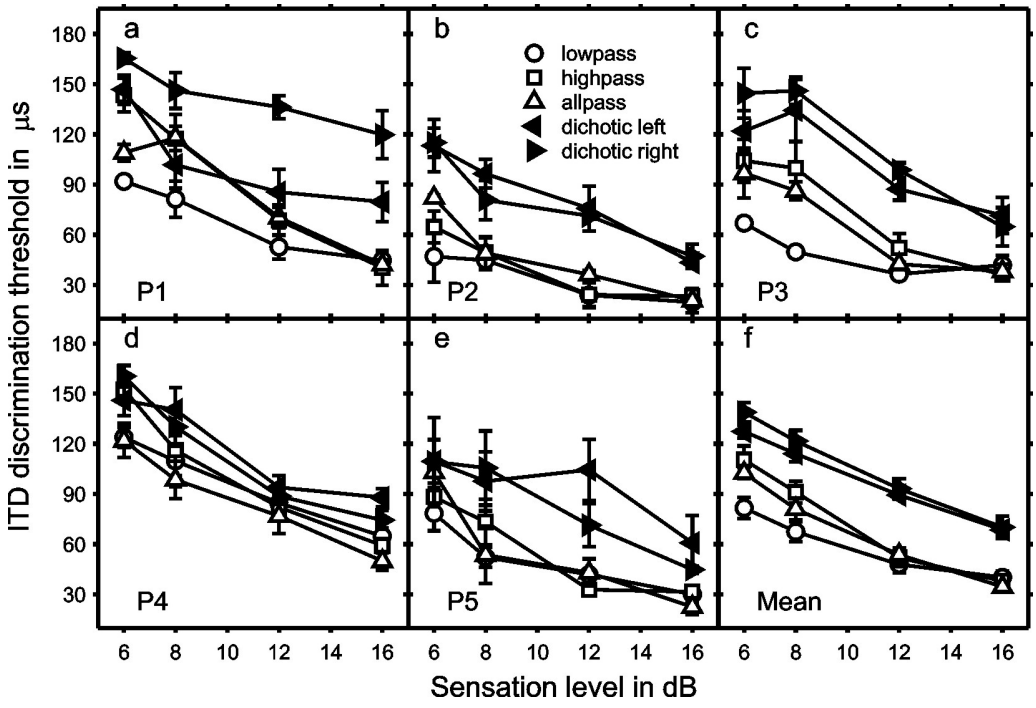

FIG. 1.2 ITD discrimination threshold for a partially masked 500-Hz pure tone plotted as a function of the tone's sensation level in dB. The parameter is the masking condition: the diotic conditions are shown by the open symbols (lowpass: circles; highpass: squares; allpass: upward-pointing triangles), and the dichotic conditions are shown by the filled triangles (dichotic left: left-pointing triangles; dichotic right: right-pointing triangles; see legend in panel b). Panels a-e show individual data of five participants; panel f shows the average thresholds. Error bars show standard errors.

Importantly, in the dichotic masking conditions, participants reported perceiving the tone as being lateralized towards the ear that received the lowpass masker, whereas the tone was perceived centrally in the diotic conditions. This pattern of lateralization would be expected if participants were extracting ITDs from disparate channels in the dichotic conditions, because the lowpass masker leaves audible the basal part of the tone's excitation pattern, the phase of which leads that of the apical part. However this pattern of lateralization may be due to differences in the shape of the residual excitation patterns of the partially masked tones. Model simulations presented in Sec. 1.D showed that the residual excitation patterns were broader in the lowpass than the highpass condition. This would mean that in the dichotic condition, a small interaural level difference (ILD) may occur within-channel (around 500 Hz, assuming a corresponding channel model) resulting in the signal being perceived as lateralized to the side with the lowpass masker. These subjective reports were confirmed and quantified by an additional ITD matching experiment, which measured the external ITD that would be necessary to compensate for the lateralization produced by the dichotic maskers (see Sec. 1.B). Figure 1.3 shows that the external ITD necessary for the tone to be perceived on the midline (referred to as the 'matching ITD') strongly depended on the side to which the lowpass and highpass maskers were presented [filled symbols: original participants, $F(1,2) = 59.514$, $p = 0.016$); open symbols in panel d: new participants, $t(4) = 5.238$, $p = 0.006$]. When the lowpass masker was presented to the left ear (dichotic-left condition, left-pointing triangles in Fig. 1.3), a large positive ITD (237 μs) was necessary to bring the dichotically masked tones to midline. The opposite was true for the dichotic-right condition; in this case, a large negative ITD (-398 μs) was necessary to centre the masked

tones (right-pointing triangles in Fig. 1.3).These results confirm participants' reports that, without an external ITD, the tones in the dichotic masking condition were perceived as being lateralized towards the side receiving the lowpass masker. There was a tendency, albeit non-significant, for the magnitude of the matching ITD to decrease towards higher sensation levels [$F(3,6) = 1.614$; $p = 0.282$]. There was no significant interaction between dichotic masking condition and sensation level.

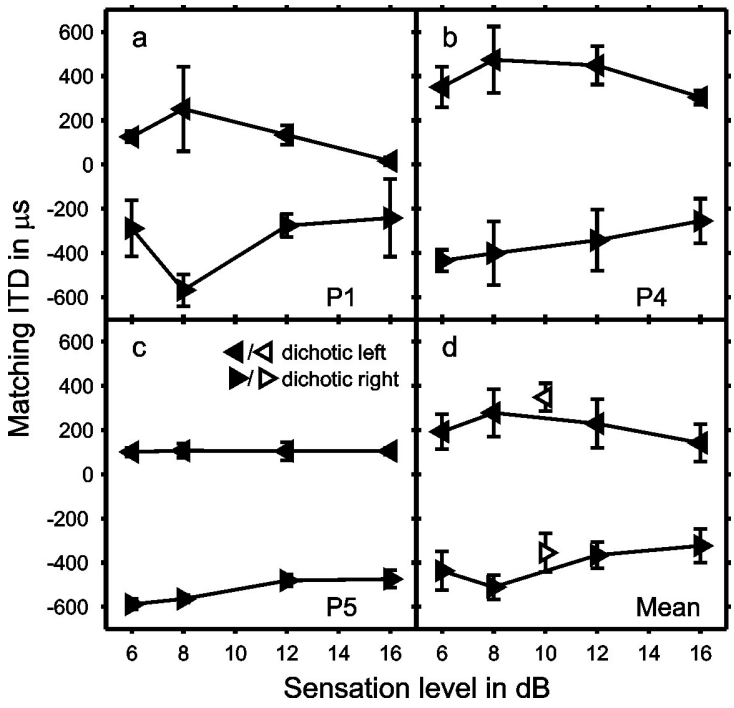

FIG. 1.3 External ITD necessary for a 500-Hz pure tone in dichotic masking conditions to be perceived on the midline ('matching ITD'). The matching ITD is plotted as a function of the tone's sensation level in dB. Dichotic-left and dichotic-right conditions are shown by left and right-pointing triangles, respectively. The filled symbols in panels a-c show individual data of three participants who had also taken part in the main experiment. The filled symbols in panel d show the average of these data. The open

symbols in panel d show the data of five new participants. Error bars show standard errors.

It is unlikely that the lateralization of the signal in the dichotic conditions explains the large ITD discrimination thresholds in these conditions, because the ITD discrimination threshold for pure tones is largely independent of ITD as long as the ITD is not ambiguous (Domnitz and Colburn, 1977; see, however, Mossop and Culling, 1998).

At first sight, the finding of larger ITD discrimination thresholds in the dichotic compared to the diotic conditions would suggest that the auditory system extracts ITDs from corresponding channels only. However, the next section shows that the data can equally be accounted for by an ITD processing model based on cross-channel comparisons, if the range of cross-channel comparisons is restricted to produce a physiologically plausible range of internal delays.

The *post hoc* tests showed a significant difference between the lowpass and highpass conditions, with the lowpass condition (open circles in Fig. 1.2) yielding smaller thresholds than the highpass condition (open squares; $p = 0.048$) at the lower two of the four sensation levels. The difference between the low- and highpass conditions in the group data was largely due to two of the five participants (P1 and P3, panels a and c in Fig. 1.2), who showed substantially smaller thresholds in the lowpass condition than in the other two diotic conditions (highpass and allpass). This difference might be related to the fact that the apical part of the basilar membrane travelling-wave response to pure tones has a steeper phase gradient than the basal part of the response (Ren, 2002; van der Heijden and Joris, 2006; for review, see Robles and Ruggero, 2001). A steeper phase gradient would be expected to be detrimental to an ITD processing mechanism that relies on the comparison between corresponding frequency channels from the two ears, because

mismatches in channel alignment, due to imprecisions in synaptic innervation of binaural neurons, would lead to larger errors in the ITD estimate, the larger the phase gradient of the cochlear response. In contrast, a steep phase gradient would be beneficial for a cross-channel ITD processing mechanism, because the phase gradient is what produces the internal delays in a cross-channel mechanism. Thus, if the small difference between the diotic lowpass and highpass conditions observed in the current data is due to the difference in phase gradient between the basal and apical parts of the cochlear travelling-wave response, this difference would argue in favour of a corresponding-channel ITD mechanism. However, the difference could also be due to differences in the shape of the residual excitation patterns of the partially masked tones between these masking conditions. Model simulations presented in Sec. 1.D showed that the residual excitation patterns were broadest in the lowpass condition (see Fig. 1.6), which may explain why this condition yielded the smallest ITD discrimination thresholds.

Individual results

While participant 4's individual data showed a significant main effect of masking condition [$F(4,12) = 4.454$, $p = 0.02$] and the general pattern of this participant's results was consistent with that seen in the average results, participant 4 showed a much smaller threshold difference between the diotic and dichotic conditions than the other participants (Fig. 1.2d). Participant 4 also had the highest audiometric thresholds at 500 Hz (about 25 dB HL). However, it is unlikely that these elevated audiometric thresholds would account for the anomalous ITD discrimination results, because participant 4's masked thresholds were very similar to the average masked thresholds. The fact that the ITD discrimination threshold of participant 4 showed a similar decrease with increasing sensation level as the

average threshold (compare Fig. 1.2f) suggests that the anomalous pattern of results was not simply due to a floor effect.

Participant 1, surprisingly, showed a substantial asymmetry between the two dichotic conditions (dichotic left and right; $p<0.001$), in that ITD discrimination performance in the dichotic-right condition (lowpass masker in right ear; see right-pointing filled triangles in Fig. 1.2a) was much worse than in the dichotic-left condition (lowpass masker in left ear; left-pointing triangles); performance in the dichotic-left condition was almost as good as in the diotic highpass and allpass conditions (open squares and upwards-pointing triangles). Such an asymmetry would be difficult to explain in an ITD model based on comparisons between corresponding channels only. In contrast, in Sec. 1.D, we show that any asymmetries between the dichotic-left and dichotic-right conditions can be readily accounted for by a stereausis-type mechanism by making assumptions about the frequency distribution of binaural connections between disparate channels.

Participant 1 had required a considerable amount of training to reliably attain a threshold estimate in the dichotic-right condition at 6 and 8 dB SL. This prompted us to investigate whether the excluded participant had shown a similar asymmetry as participant 1 and whether the asymmetry had contributed to the exclusion. For that, we calculated the ratio of failed (terminated; see Sec. 1.B) to successful threshold runs for the dichotic-left and dichotic-right conditions at the lower two sensation levels (6 and 8 dB SL). The difference between these failure-to-success ratios (FSRs) for the dichotic-right and dichotic-left conditions (right–left FSR) is shown in Fig. 1.4 (open bars, right ordinate). The figure also shows the difference between the ITD discrimination

thresholds for the right and left dichotic conditions as a percentage of their mean (averaged across the lower two sensation levels) for the five participants who completed the experiment (participants 1-5; filled bars, left ordinate). The figure shows that the excluded participant (P6) indeed showed a large asymmetry in the FSR between the left and right dichotic conditions; this participant successfully completed only 6 of a total of 48 threshold runs in the dichotic left condition, compared to 13 out of 18 successful runs in the dichotic right condition. Note that the excluded participant's asymmetry was even larger than and opposite to that observed in participant 1 and that the asymmetry was effectively the reason for exclusion. The presence of asymmetry did not seem to be related to the degree of handedness; according to the Edinburgh inventory (Oldfield, 1971), participants 1 and 6, who showed an asymmetry between the dichotic conditions, exhibited the highest (100) and lowest (66) laterality ratios of the group, respectively.

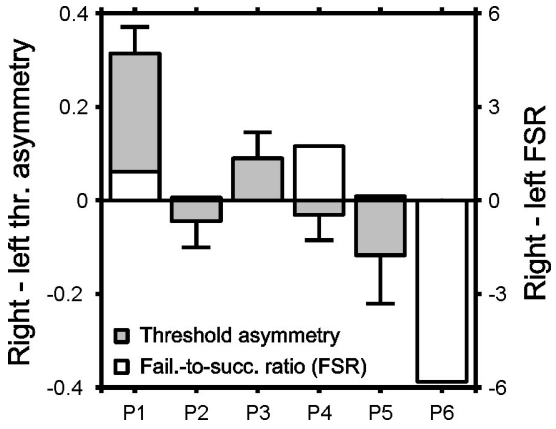

FIG. 1.4 Asymmetry in ITD discrimination performance between dichotic-right and dichotic-left conditions. The grey bars show the difference in ITD discrimination threshold between the dichotic-right and dichotic-left conditions, expressed as a proportion of their mean and averaged across the lower two sensations levels (left ordinate), for each of the five participants who completed the ITD discrimination experiment (P1-P5). The open bars show the difference in the failure-to-success ratio (FSR; see Sec. 1.C) between the dichotic-right and dichotic-left conditions (right ordinate) for each of six participants, one of whom (P6) did not complete the experiment (see Sec. 1.C).

1.D MODEL SIMULATIONS

The previous section showed that the ITD discrimination threshold for a partially masked 500-Hz pure tone was on average more than 50% larger for the spectrally dichotic (dichotic left and right) than for the diotic maskers (lowpass, highpass and allpass). The difference between the dichotic and diotic masking conditions was largely independent of the tone sensation level over the range of sensation levels tested (6-16 dB SL). In this section, the measured ITD discrimination thresholds are compared to predictions from two opposing models of ITD processing, one representing ITD processing based on interaural temporal comparisons between corresponding frequency channels from the two ears by means of neurally generated internal delays, and the other representing processing based on comparisons between non-corresponding channels with internal delays being generated mechanically at the level of the cochlea (see Sec. 1.A). The first model was a computational version of Jeffress' (1948) delay-and-coincidence mechanism, in which the cross-correlation function between corresponding frequency channels was computed for a range of correlation lags, assumed to be created by axonal propagation delays (APDs). This model will henceforth be referred to as the APD model. The second model was a computational version of Schroeder's (1977) cross-channel mechanism, in which the cross-correlation at lag zero was calculated between all possible channel combinations within a range of frequencies around the signal frequency (500 Hz). This model will henceforth be referred to as the cross-channel (XCH) model.

Model architecture

Both models (APD and XCH) consisted of four stages: the first and second stages simulated the peripheral response to the pure-tone signal and the effect of the noise

masker, the third stage simulated the binaural processing of the signals from the left and right ears, and the last stage calculated a decision measure upon which the threshold estimates were based.

Stage 1. Peripheral response to pure-tone signal

The peripheral response to the pure-tone signal was modelled using the DSAM/AMS software package (O'Mard and Meddis, 2004). The spectral analysis performed by the cochlea was simulated by a 65-channel gammatone filterbank with centre frequencies between 200 Hz and 1 kHz (±5 ERBs around 500 Hz), evenly distributed on an ERB scale (6.5 channels per ERB). The filterbank output was halfwave-rectified and fourth-order lowpass-filtered at 1 kHz to simulate the transformation from the mechanical response of the basilar membrane to the neural activity pattern (NAP) flowing up the auditory nerve. The first 50 ms of the NAP were discarded to allow the model response to reach a steady state. Finally, each channel output was normalized by its root-mean-square (rms) amplitude.

Stage 2. Effect of noise masker

The effect of the noise masker was modelled by first simulating the detection thresholds using Glasberg and Moore's (1990) excitation-pattern model. The detection thresholds were simulated by calculating the positive difference, $R = \max(NS\text{-}N, 0)$, between the excitation pattern for the tone signal plus noise (NS; expressed in dB per ERB) and for the noise alone (N); R is referred to as the residual excitation pattern. The noise, N, was the sum of the external noise and an internal noise, N_0, which had a constant level per ERB. The average tone detection threshold in quiet (dotted horizontal line in Fig. 1.1) was used to calculate the internal noise level, N_0. All excitation patterns

were calculated with 512 channels between 20 Hz and 1.9 kHz (±10 ERBs around 500 Hz), evenly distributed on an ERB scale (25.6 channels per ERB). Model detection threshold was defined as the tone level at which the residual excitation pattern reached a criterion area, A_{crit}, which was a free parameter in the fitting process. All conditions of the detection threshold measurements were fitted simultaneously, with the same value of A_{crit}, which was varied to minimize the rms deviation between the simulated and observed thresholds. Figure 1.5 shows that the model (black symbols and lines) produced a reasonably good fit to the detection threshold data (grey symbols and lines). As would be expected based on the linear nature of the model, the model was unable to simulate the difference between the lowpass and highpass conditions at masker cutoff frequencies close to the signal frequency.

The simulated detection thresholds were then used to calculate the cutoff frequencies for the lowpass and highpass maskers, and the levels for the allpass masker to yield the appropriate tone sensation levels for the simulation of the ITD discrimination thresholds (6, 8, 12 and 16 dB SL); this was done in exactly the same was as in the experiment (see Sec. 1.B). Finally, residual tone excitation patterns were calculated for all masker types (lowpass, highpass and allpass) and sensation levels (6, 8, 12, and 16 dB SL). The residual excitation patterns were converted to linear amplitude units, normalized to the maximum of the pattern with the highest peak (i.e., the pattern for the highest sensation level in the allpass condition; see Fig. 1.6) and interpolated to the channel frequencies of the gammatone filterbank used in the first stage of the simulation. To account for the fact that the ITD discrimination threshold functions reach an asymptote at high sensation levels, the residual excitation patterns were compressed by exponentiation

with a parameter $\alpha < 1$, which was a free parameter in the ITD discrimination threshold simulation. Compression makes the excitation patterns for different sensation levels more similar. The normalization of the patterns to a maximum value of unity meant that the compression did not change the codomain of the patterns. The patterns shown in Fig. 1.6 are uncompressed.

Stage 3. Binaural processing

In the next stage, the normalized NAPs from each ear were weighted by the appropriate residual excitation pattern from Stage 2 and then combined to extract the interaural temporal information contained in the resulting patterns.

For the APD model, each channel from the left ear was cross-correlated with the corresponding channel from the right ear over a range of lags between ±1280 μs by integrating the cross-product between the non-mean-corrected channel waveforms (see Bernstein and Trahiotis, 1996) over time. Panels a and b in Fig. 1.7 show the cross-correlation patterns of the normalized NAPs for 500-Hz tones leading by 190 μs at the left (a) or right ear (b); to show the entire pattern, the NAPs were not yet weighted by the residual excitation patterns in this example. The figure reveals a vertical ridge of high correlation, the horizontal position of which reflects the ITD of the tone (white-shaded areas in Fig. 1.7a,b).

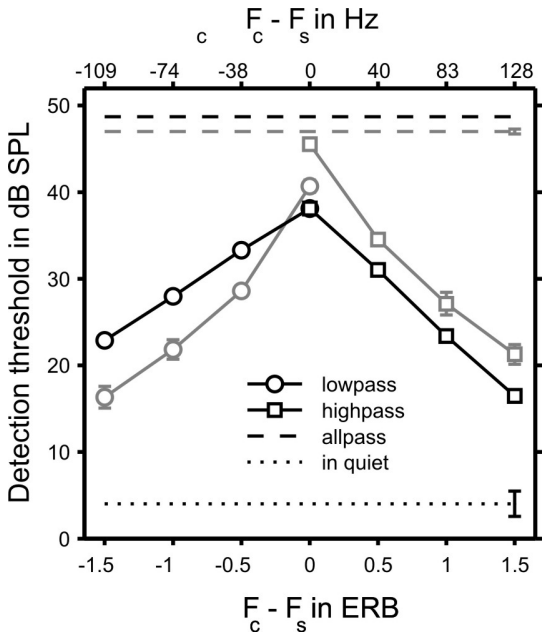

FIG. 1.5 Model simulations of the detection threshold data from Fig. 1.1. The simulated thresholds are shown by the black symbols and lines and plotted in the same way as the measured thresholds in Fig. 1.1. The measured thresholds are replotted for comparison (grey symbols and lines). The parameter is the masking condition (see legend).

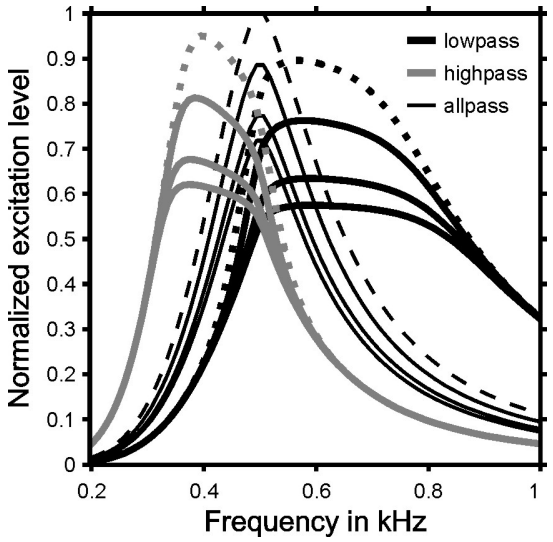

FIG. 1.6 Simulated residual excitation patterns of a 500-Hz pure tone in different masking conditions and for different sensation levels (bold black lines: lowpass; bold grey lines: highpass; thin black lines: allpass). Excitation level was expressed in linear units and normalized to the maximum of the pattern with the highest peak, i.e., the pattern for the highest sensation level in the allpass condition (dashed thin black line). The dotted bold black and grey lines show the patterns for the lowpass and highpass conditions, respectively, for the highest sensation level; these patterns were used to create the cross-correlation patterns shown in Figs 1.7d and 1.8d (see Sec. 1.D).

For the XCH model, each channel in the left-ear NAP was correlated (lag zero) with all channels in the right-ear NAP, again using the cross-product between the non-mean-corrected waveforms. Panels a and b in Fig. 1.8 show that the ridge of high correlation along the central diagonal (marked by a dashed line) in the cross-channel correlation pattern of a 500-Hz tone is shifted towards the upper left corner of the pattern when the tone is leading in the left ear (by 190 μs as in Fig. 1.7) and towards the lower right corner when the tone is leading in the right ear. The central diagonal represents points where the left- and right-ear channels have the same frequencies. Panel c in Figs 1.7 and 1.8 show the difference between the cross-correlation patterns for the right-leading and left-leading tones shown in panels b and a. These difference patterns simulate the information potentially available to the binaural system in the 2I2AFC paradigm used in the ITD discrimination threshold measurements. The effect of the masking noise would be to obscure part of these patterns. Panel d in Figs 1.7 and 1.8 shows the difference patterns from panel c after weighting with the residual excitation patterns. The difference pattern in Fig. 1.7d was weighted with the residual excitation pattern for the diotic lowpass condition at 16 dB SL (see bold black dotted line in Fig. 1.6); the figure shows that this masking condition obscures the low-frequency part of the difference pattern. In Fig. 1.8d, the difference pattern was weighted with the residual excitation pattern for the 16-dB SL dichotic-left condition (see bold black and grey dotted lines in Fig. 1.6), which obscures the low frequencies in the left ear and the high frequencies in the right ear and thus limits the difference pattern to the lower right quadrant.

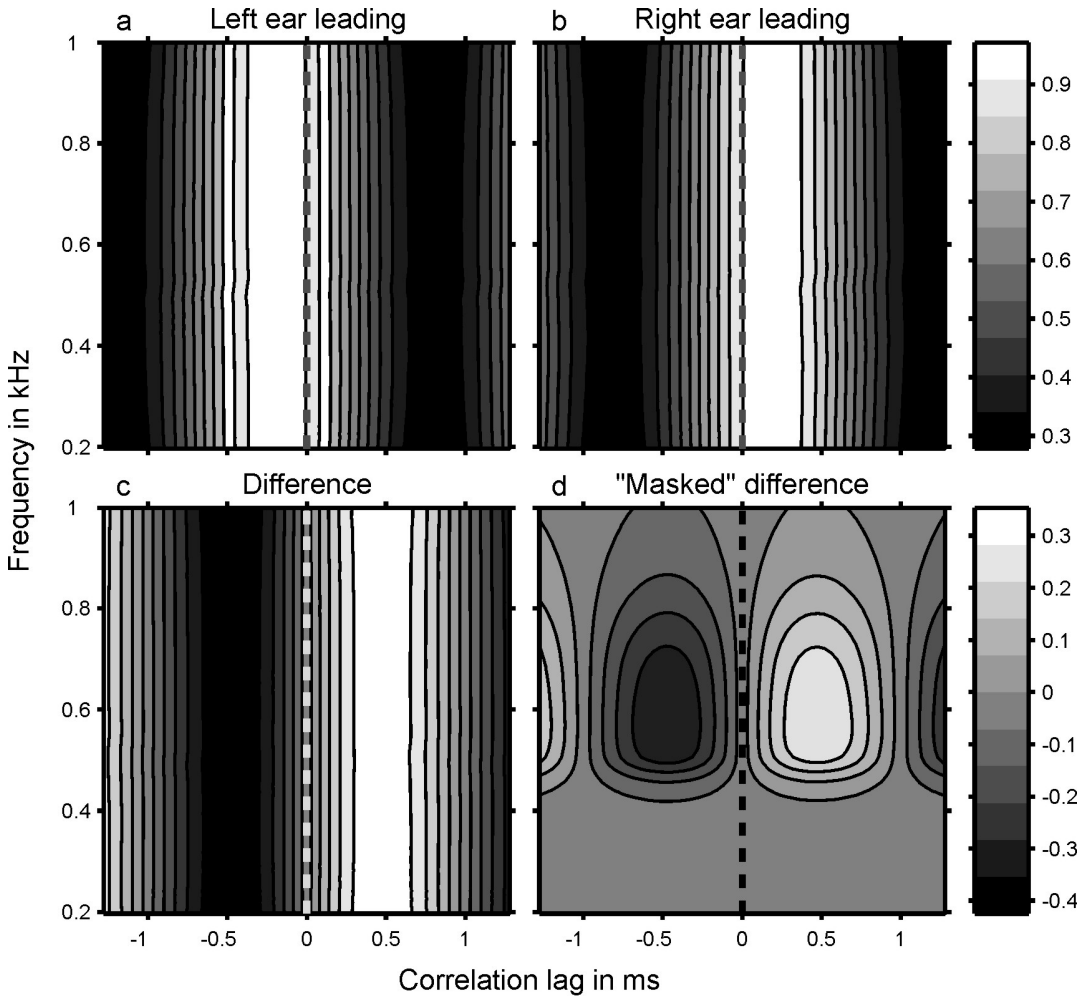

FIG. 1.7 Axonal propagation delay (APD) model. Panels a and b show the cross-correlation patterns of the normalized NAPs for a 500-Hz tone leading by 190 μs at the left and right ear, respectively; different correlation values are represented by different grey shades (see colour bar to the right of panel b). Panel c shows the difference between the patterns in panels a and b (b - a) and panel d shows the difference pattern after weighting with the residual excitation patterns for the diotic lowpass condition at 16 dB SL (shown by dotted line in Fig. 1.6); again different values of the difference are represented by different grey shades (see colour bar to the right of panel d). The dashed vertical lines mark the point at which the internal (cross-correlation) delay is zero.

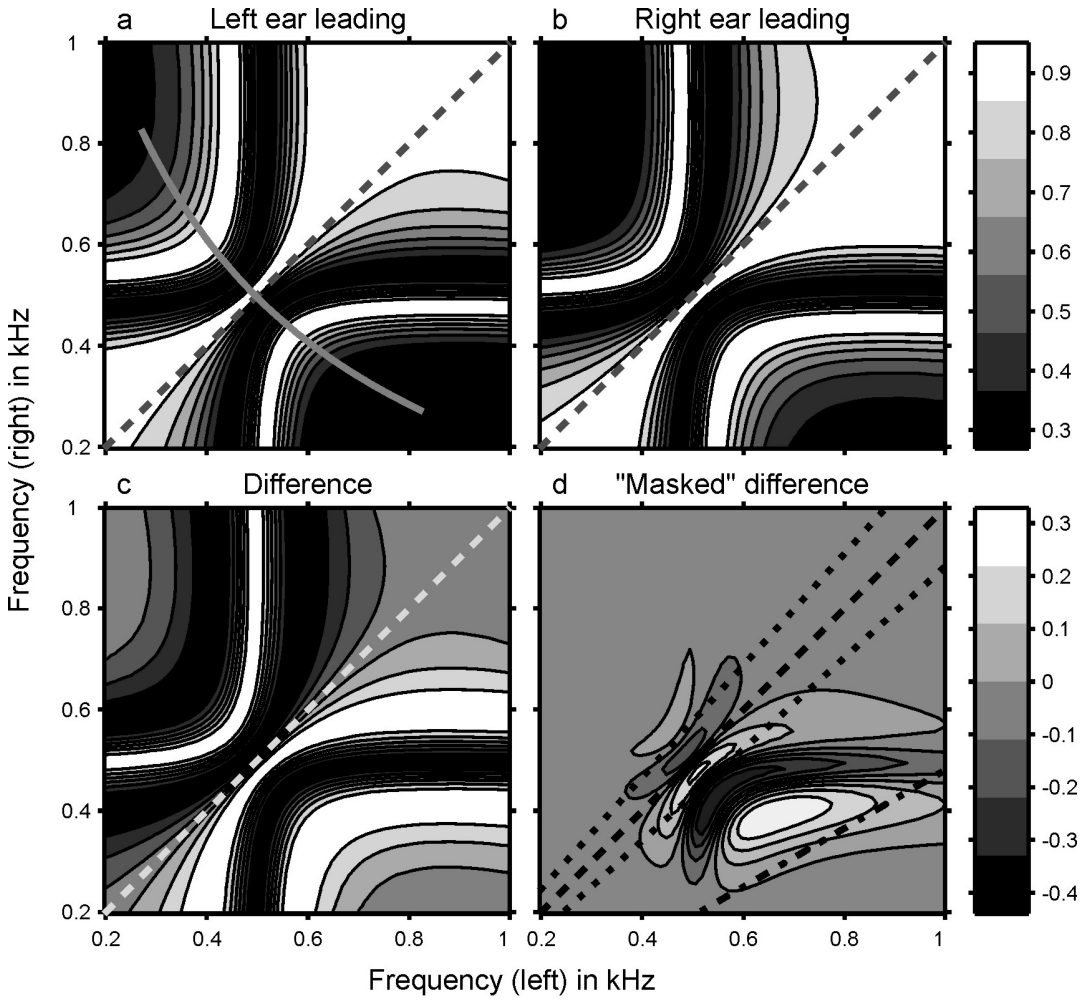

FIG. 1.8 Cross-channel (XCH) model, plotted in the same way as the APD model in Fig. 1.7. Panels a and b show the cross-correlation patterns for the same left- and right-leading tones as used in Fig. 1.7, panel c shows the difference of the patterns in panels a and b (b – a, as in Fig. 1.7), and panel d shows the difference pattern after weighting with the residual excitation patterns for the dichotic-left condition at 16 dB SL (see dotted bold black and grey lines in Fig. 1.6). The dashed diagonal lines represent points where the channel frequencies from the left and right ears are equal. The dotted lines to the left and right of the diagonal in panel d represent the boundaries of the frequency band within which the phase differences between the left- and right-ear channels range between $\pm\pi$

radians of their respective average frequency (in ERBs; see also Fig. 1.10). The dash-dotted line shows the outer boundary of one of the two slip-cycle bands, where the phase differences between the left- and right-ear channels range from π to 3π. The grey curved line in panel a shows points covered by the cross section of the cross-correlation pattern shown in Fig. 1.10.

Stage 4. Decision measure

For a given value of the compression exponent, α, difference patterns as shown in Figs 1.7d and 1.8d were calculated for all masking conditions and sensation levels, using 10 different ITDs, equally spaced between 10 and 190 μs (Figs 1.7d and 1.8d show examples for the 190-μs ITD). The decision measure, *D*, was the square root of the integral of the squared difference patterns. Threshold was defined as the ITD at which *D* reached a criterion value, D_{crit}, which was a free parameter of the fitting process and was chosen to minimize the rms deviation between predicted and observed thresholds. The other free parameter was the compression exponent, α. The value of α was determined by repeating the fitting process for each of 31 values of α, equally spaced between 0.1 and 0.25, and choosing the value that minimized the rms deviation between predicted and observed thresholds. Smaller values of α (more compression) made the slope of the function relating the simulated ITD thresholds to the tone sensation levels shallower.

Simulation results and interim discussion

The black symbols and lines in Fig. 1.9a show that the APD model produced a remarkably good fit to the experimental data (grey symbols and lines); the rms deviation between the simulated and measured thresholds amounted to only 8.5 μs. The best-fitting compression exponent, α, was 0.19 in this simulation. The model yielded similar thresholds for all three diotic masking conditions (lowpass, highpass and allpass; open symbols) and correctly predicted larger thresholds for the dichotic conditions (filled symbols). The difference between the simulated dichotic and diotic thresholds was similar to that in the data. In contrast, the XCH model provided a poor fit to the data (rms deviation = 24.6 μs), predicting largely similar thresholds for all conditions (Fig. 1.9b),

with the smallest thresholds for the dichotic and the lowpass conditions (filled triangles and open circles) and the largest thresholds for the diotic allpass condition (open triangles). The best-fitting α amounted to 0.22 in the XCH simulation. Note that, while the corresponding-channel model used in the current simulations was based on Jeffress' (1948) delay-and-coincidence mechanism and thus involved axonal propagation delays (APDs), similar findings would also be expected to apply to other models, where ITDs are extracted from corresponding channels, but the ITD analysis is based on another mechanism, such as inhibition (Brand et al., 2002; for review, see Grothe, 2003, McAlpine and Grothe 2003 and Joris and Yin, 2007).

At first glance, these modelling results seem to provide strong evidence for a mechanism based on comparisons between corresponding frequency channels. However, a possible reason for the poor performance of the XCH model may have been that the model included more cross-channel information than necessary, which unduly benefited the dichotic conditions. The cross-channel correlation patterns of the XCH model contained a relatively high degree of informational redundancy, in that, in addition to the correlation peak near the diagonal (white shading near dashed diagonal line in Fig. 1.8a,b), the patterns contained another correlation peak (representing the first slip cycle) on either side of the diagonal (hyperbolically-shaped white-shaded areas). This is illustrated in Fig. 1.10, which shows a slice of the cross-channel correlation pattern for a tone with zero ITD taken along the curved grey line in Fig. 1.8a; this line represents points where the left- and right-ear channel frequencies have equal separation (in ERBs) from 500 Hz, where the line crosses the diagonal.

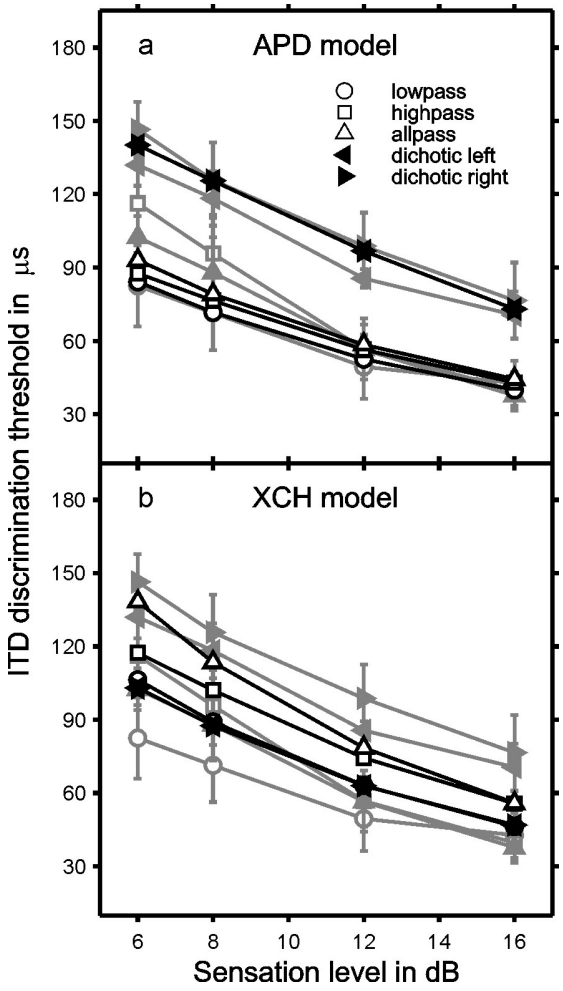

FIG. 1.9 Simulated ITD discrimination thresholds (black symbols and lines) based on the APD (a) and XCH (b) models, plotted in the same way as the measured thresholds in Fig. 1.2. The parameter is the masking condition (see legend). The average measured thresholds were replotted from Fig. 1.2f for comparison (grey symbols and lines).

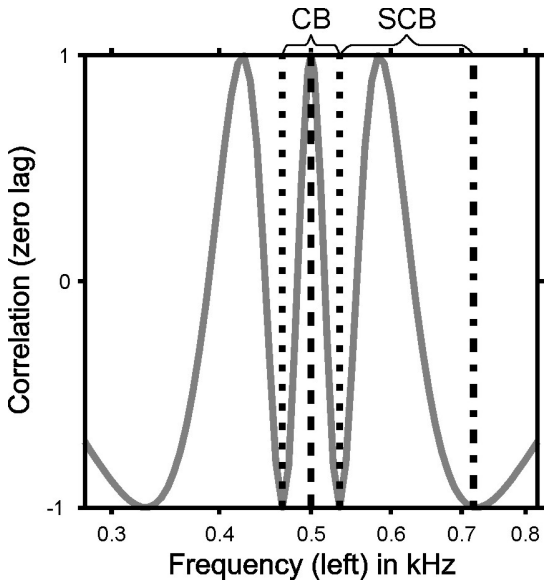

FIG. 1.10 Cross section through the cross-correlation pattern for a 500-Hz tone with zero ITD taken along the curved grey line in Fig. 1.8a, which represents points where the left- and right-ear channel frequencies have equal separation (in ERBs) from 500 Hz, which is where the line crosses the diagonal. The correlation values are plotted as a function of the channel frequencies in the left ear. As in Fig. 1.8d, the dashed vertical line represents the point where the channel frequencies from the left and right ears are equal (500 Hz). The dotted lines show the boundaries of the frequency band within which the phase differences between the left- and right-ear channels range between $\pm\pi$ (referred to as the central band, CB) and the dash-dotted line marks the outer boundary of one of the slip-cycle bands (SCB).

The degree of redundancy in the XCH patterns was (largely arbitrarily) determined by the frequency range of input channels (filter frequencies in the gammatone filterbank) and may be argued to be physiologically implausible, given that the best ITDs of binaural neurons in the mammalian brainstem seem to be limited to the range of $\pm\pi$ radians of the neurons' best frequency (which excludes the slip cycles; McAlpine et al., 1996). The XCH simulation was therefore repeated, this time limiting the cross-correlation patterns to a physiologically more plausible range. For that, the XCH patterns were limited to a band around the diagonal, shown by the dotted lines to the left and right of the diagonal in Fig. 1.8d, and rerunning the simulation as before with the limited patterns. Within the band, the phase differences between the left- and right-ear channels ranged between $\pm\pi$ radians of their respective average frequency (in ERBs; see vertical dotted lines in Fig. 1.10). The band was about 1.2 ERBs wide. Figure 1.11 shows that the limited XCH model indeed produced a better fit to the data. Like the APD model, the limited XCH model yielded the largest thresholds for the dichotic conditions and the difference between the simulated thresholds for the dichotic and the diotic conditions was similar to that in the data. The rms deviation between predicted and measured thresholds amounted to only 10.4 μs, which is comparable to that for the APD model. The best-fitting compression exponent, α, for the limited XCH simulation was similar to that for the original XCH simulation (0.215).

Accounting for the asymmetries between the dichotic conditions in individual data sets

By limiting the frequency range of cross-channel comparisons, the XCH model could be made to provide as good an account of the dichotic ITD discrimination thresholds as the APD model. In a similar way, the XCH model can also be made to

explain another aspect of the data, namely the asymmetry between the thresholds for the dichotic left and dichotic right conditions observed in some participants (see Sec. 1.C), by limiting the cross-channel correlation patterns to an asymmetrical band around the central diagonal. In particular, a band that includes the slip cycle on the right of the diagonal (higher frequencies in the left ear, lower frequencies in the right ear) would be expected to predict better performance for the dichotic-left (lowpass masker in the left ear) than the dichotic-right condition, whereas a band that includes the left slip cycle would be expected to favour the dichotic-right condition. To test this, the data of participant 1, whose thresholds for the dichotic-left condition were substantially smaller than those for the dichotic-right condition (Fig. 1.2a) were fitted by limiting the XCH patterns to a band that included the slip cycle on the right but not the left of the diagonal (see dash-dotted line in Figs 1.8d and 1.10). The relative weight of the central and the slip-cycle bands [denoted CB (central band) and SCB (slip cycle band) in Fig. 1.10] was a free parameter in this simulation and was adjusted to minimize the rms deviation between the simulated and participant 1's individual thresholds.

Figure 1.12 shows that this asymmetrically weighted XCH model provided a reasonably good account of the asymmetry between the dichotic conditions in participant 1's ITD threshold data. The relative weights of the central and slip-cycle bands that minimized the rms deviation between simulated and measured thresholds (17.9 μs) amounted to 70% and 30%, respectively. The compression exponent, α, amounted to 0.185 in this simulation. It should be noted that the XCH model could similarly be made to explain the individual results of participant 4, who showed only marginal differences between the ITD discrimination thresholds for the dichotic and diotic conditions (Fig.

1.2d). Individual differences and asymmetries in the frequency distribution of cross-channel connections in the XCH model could arise as a consequence of incomplete pruning during development. It would seem difficult or impossible to see how any model that is purely based on comparisons between corresponding frequency channels (like the APD model), could explain the asymmetry in the effect of combining activity in different frequency channels shown by participant 1, unless comparisons between non-corresponding channels were also included in the model.

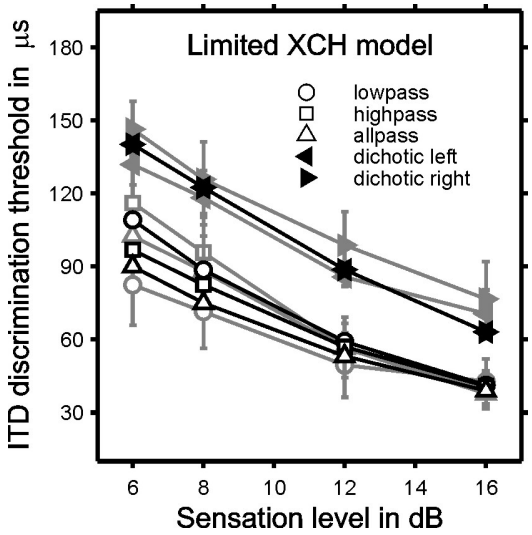

FIG. 1.11 Simulated ITD discrimination thresholds (black symbols and lines) based on the limited XCH model. As in Fig. 1.9, the grey symbols and lines show the average measured thresholds for comparison.

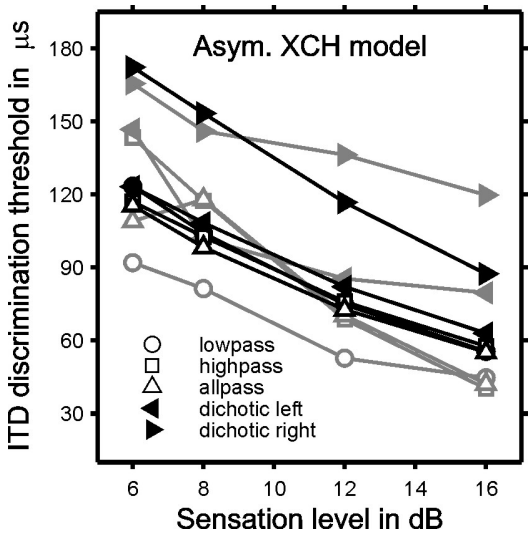

FIG. 1.12 Simulated ITD discrimination threshold for participant 1 (black symbols and lines) based on the asymmetrically weighted XCH model. Participants 1's measured thresholds were replotted from Fig. 1.2a for comparison (grey symbols and lines).

1.E SUMMARY AND CONCLUSIONS

The current study showed that the ITD discrimination threshold for pure tones partially masked by high- or lowpass noise maskers was about twice as large in the dichotic masking conditions, where the masker spectral characteristics differed between the two ears, than in the diotic conditions, where the masker had the same spectral properties in both ears. The dichotic conditions were intended to force the participants to extract interaural temporal information from disparate frequency channels in the two ears. The finding that ITD discrimination performance was poor in these conditions at first sight suggested that the auditory system extracts interaural temporal information mainly from corresponding frequency channels in the two ears. However, the model simulations showed that a computational version of Schroeder's (1977) cross-channel (XCH) model of ITD processing was able to provide a similarly accurate account of the current data as a corresponding-channel model based on Jeffress' (1948) delay-and-coincidence mechanism (APD model), if the range of cross-channel comparisons was restricted so as to produce a physiologically plausible range of internal delays. The XCH model was also able to explain the asymmetry observed in the dichotic ITD discrimination thresholds of some participants, which would have been difficult, if not impossible, to explain with the APD model, unless cross-channel comparisons were also included in the APD model.

The fact that, in the dichotic masking conditions, participants perceived the tone as being lateralized towards the ear receiving the lowpass masker, suggests that participants may indeed have been using information from disparate channels in these conditions. The model results suggest that, if performance in the dichotic conditions was based on cross-channel comparisons, such comparisons would have to be limited to a

fairly narrow frequency range of little more than one auditory-filter bandwidth around each channel. The modelling also showed that even within this narrow range, phase differences between non-corresponding channels cover the entire range of internal delays observed in ITD-sensitive neurons.

However, whether or not these phase delays actually contribute to the internal delays for fine-structure ITD processing cannot be determined from the current data. Conceivably, cross-channel comparisons could also be part of a mechanism based on neurally generated internal delays to make the mechanism robust against channel mismatch. Alternatively, neurally generated delays and cochlear phase delays may both contribute to ITD sensitivity at low frequencies.

Future studies could use partial masking to investigate the effect of sound level, which has been shown to affect cochlea phase delays, on the lateralization of partially-masked tones.

Chapter 2. Evidence for opponent-channel coding of interaural temporal cues to sound lateralisation in human auditory cortex[2]

2.A INTRODUCTION

As described in **Chapter 1** (Sec. 1.A), horizontal sound localisation in humans is mainly based on differences in sound arrival time at the two ears. Humans are sensitive to interaural time differences (ITDs) of only a few tens of microseconds (Klumpp and Eady, 1956). The initial processing of these minute time differences involves brainstem structures that are highly specialised in temporal processing (for review see Oertel, 1999; Grothe, 2003; Joris and Yin, 2007). Traditionally, it was assumed that these structures convert ITDs into a topographic (or rate-place) representation, based on an array of neurons tuned to different ITDs (Fig. 2.1a; Jeffress, 1948; see Sec. 1.A); the incoming ITD is assumed to be inferred from the maximum or centroid of the distribution of activity across the array. While the topographic model seems to be appropriate for owls (Sullivan and Konishi, 1986; Carr and Konishi, 1990; see Konishi, 2003, for review) and chickens (Overholt et al., 1992), physiological results suggest that it may not be generally applicable to mammals. These results have shown that, in mammals with a small head and thus a small range of naturally occurring ITDs (referred to as the "physiological range"), most ITD-sensitive neurons respond best to ITDs outside of that range (McAlpine, 1996, 2001; Fitzpatrick et al., 2000; see also Sec. 1.A). These studies have also shown that the largest ITDs are limited to about half the period of the relevant neuron's best frequency (McAlpine, 1996; see also Thompson et al., 2006). This has led to the suggestion that, in these species, ITDs may be coded by a non-topographic population rate code, involving only two hemispheric channels, broadly tuned to ITDs

[2] Based on Magezi and Krumbholz (2009b)

from the contralateral hemifield. In this hemispheric-channel model, each channel's overall activity level is assumed to increase with increasing contralateral ITD and the incoming ITD is assumed to be inferred from the relative balance of activity between the two channels (Fig. 2.1b; von Békésy, 1930; van Bergeijk, 1962; Colburn and Latimer, 1978; McAlpine et al., 2001; Harper and McAlpine, 2004; see McAlpine, 2005, for review).

It remains unclear, which of the two models applies to humans. A recent modelling study suggests that, if the distribution of the best ITDs of ITD-sensitive neurons were freely adaptable to ecological constraints, the ITD coding mechanism pertaining to a given species would depend on the species' head size and the spectral range over which the ITD information is used (Harper and McAlpine, 2004). In that case, ITD processing in humans would be expected to be based on a topographic or intermediate code for all but the lowest frequencies (< 250 Hz). Alternatively, the ITD coding mechanism may be a characteristic of the phylogenetic class that a given species belongs to (e.g., mammals versus birds; see McAlpine and Grothe, 2003, for review), possibly determined by the physiological mechanism by which the internal delays in the input to ITD-sensitive neurons are generated (see, e.g., Brand et al., 2002; Joris et al, 2006). In that case, ITD processing in humans may be expected to be based on a hemispheric-channel code.

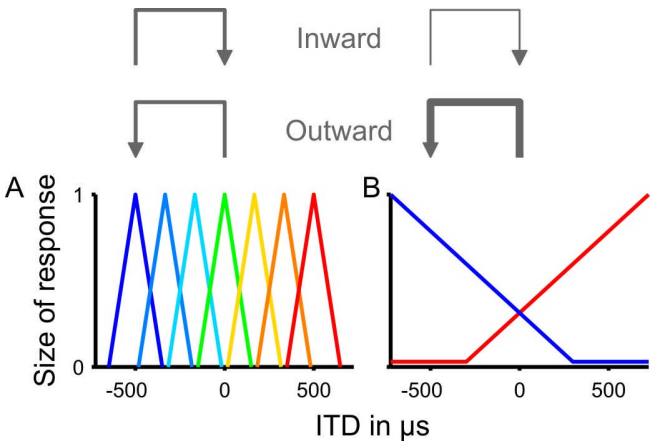

FIG. 2.1 Schematic representation and predictions of the topographic (a) and hemispheric-channel models (b). The coloured lines show the tuning characteristics of different ITD channels as a function of ITD. The arrows at the top show ITD changes away from and towards the midline. The line thickness signifies the relative amplitude of the corresponding ITD change responses predicted by the respective model.

The current study investigates this question with electroencephalography (EEG) and a specialised stimulation paradigm, which makes it possible to isolate the response of those neural elements in auditory cortex specifically involved in ITD processing (Ungan et al., 1989, 2001; Jones et al., 1991). The paradigm is referred to as the "continuous stimulation paradigm" (CSP; Hewson-Stoate et al., 2006) and involves preceding a test sound (black in Fig. 2.2a) with a control sound (grey) differing only in ITD, and measuring the response to the transition from the control to the test sound. If the control sound is long enough to allow the transient response to its onset to subside before the onset of the test sound, the transition response would be assumed to reflect activity only from those neural elements that are more strongly activated by the test than the control sound. In that case, the size of the transition response for a given ITD change between the control and test sounds would be expected to depend on the mechanism by which ITDs are coded in auditory cortex. According to the topographic model, the amplitude of the transition response should be mainly determined by the size of the ITD change and be little influenced by its direction. Thus, the response to an ITD change towards the midline ("inward" change) should have the same or similar amplitude as the response to the reverse change away from the midline ("outward" change; see arrows in Fig. 2.1a). If at all, the inward response may be expected to be slightly larger than the outward response, if a greater density of neurons tuned to ITDs near zero is assumed (Colburn, 1973; Stern and Shear, 1996). In contrast, the hemispheric-channel model would predict the response to an outward ITD change to be much larger than the response to the corresponding inward change (see arrows in Fig. 2.1b). In fact, the contralateral channel would be expected to produce little or no response to the inward change at all, and so, the inward

response would be expected to reflect whatever little activity the test ITD elicits in the ipsilateral channel. The current experiment was designed to test these opposing predictions.

2.B METHODS

Stimuli

The stimuli consisted of a 1,500-ms control portion and a 250-ms test portion and were presented with an inter-stimulus interval of 1,500 ms. Both the control and test portions consisted of random noise (Fig. 2.2a), generated afresh for each trial. The ITD was fixed throughout the control portion and changed to a different static value for the test portion, creating the perception of an abrupt shift in the intracranial position of the stimulus at the transition. Six different stimulus conditions were tested. In the "outward" conditions, the ITD change was from a more medial to a more lateral position (away from the midline). There were three outward conditions with ITDs changing from 0 (midline) to -250 (lateralised about halfway to the left ear; Toole and Sayers, 1965a) or -500 μs (practically fully lateralised towards the left ear), and from -250 to -500 μs (upper arrows in Fig. 2.3). In the other three conditions, referred to as "inward" conditions, the ITD changes were reversed relative to the outward conditions (-250 to 0 μs, -500 to 0 μs and -500 to -250 μs; lower arrows in Fig. 2.3). The ITD changes were limited to the left hemifield, because previous neuroimaging and electrophysiological studies have shown that, in humans, the response to left-lateralised sounds is more contralateral, similar to animals, and is often larger than the response to right-lateralised sounds (see Krumbholz et al. 2007 for a review of recent imaging data in humans, and Malhotra et al., 2004, 2008, for animal data). The hemispheric distributions of the responses were expected to

provide potential cues for distinguishing between different ITD processing mechanisms. In order to ensure that all spectral components of the current stimuli would convey unambiguous ITD cues (i.e., component interaural phase differences were never greater than π), stimuli were lowpass filtered at 1 kHz using a 16^{th}-order Butterworth IIR filter. Filtering was carried out after introducing the ITD change to avoid audible clicks at the transition from the control to the test portion. The stimuli were gated on and off with 5-ms cosine ramps, which were synchronous at two ears to avoid envelope ITD cues. They were presented at an overall level of about 70 dB SPL.

As in **Chapter 1** (Sec. 1.B), stimuli were generated digitally with a 25-kHz sampling rate using TDT System 3 (Tucker Davis Technologies, Alachua, FL) and MATLAB® (The Mathworks, Natick, MA). They were digital-to-analogue converted with a 24-bit amplitude resolution (TDT RP2.1), amplified (TDT HB7) and presented over headphones (K240 DF, AKG, Vienna, Austria) to the participant, who was seated in a double-walled sound-attenuated room. The experiment was divided into four runs of approximately 20 minutes each, with short breaks in between. Each of the six conditions was presented a total of 248 times (62 times within each run). Conditions were presented in a random order within each run.

Data acquisition

Auditory-evoked potentials were recorded from 33 mostly equidistant 10-20 positions using Ag/AgCl sintered ring electrodes (Easy Cap, Herrsching, Germany) and a BrainAmp DC EEG amplifier (Brain Products, Munich, Germany). The ground electrode was placed centrally on the forehead (AFz position) and the vertex channel (Cz) was used as recording reference and reconstructed by re-referencing to average reference post

recording. Data were recorded continuously with a sampling rate of 500 Hz and analogue-filtered between 0.1 and 250 Hz. Participants watched a self-chosen silent movie to maintain wakefulness during recording.

Data Analysis

Pre-processing of the raw data was performed with the EEGLAB toolbox (Delorme and Makeig, 2004), which runs under MATLAB®. The data were (i) lowpass filtered at 35 Hz using a 32nd-order zero-phase Butterworth IIR filter, (ii) re-sampled at 250 Hz to reduce computation time, (iii) re-referenced to average reference, and (iv) divided into stimulus-locked epochs covering the period from -250 ms to 3,000 ms relative to stimulus onset. Epochs with non-stereotypic artefacts were then rejected automatically using the joint probability function in EEGLAB, which identifies artefacts by looking for unusually large potentials across many electrodes. About 12% of epochs were rejected on average by this method. Stereotypic artefacts (electro-ocular and electro-cardiac activity), were eliminated by applying an independent components analysis using the extended infomax algorithm (Bell and Sejnowski 1995; Lee et al., 1999) to the remaining epochs and manually rejecting artifactual components based on inspection of the components' activity time courses, field maps and event-related average waveforms. The corrected data were back-projected and baseline-corrected to the 200-ms period preceding stimulus onset. The response to the control sound shows that the stimuli produced a sustained response (SR in Fig. 2.2b), upon which the transient response to the ITD change at the onset of the test portion (labelled "change response", or CR, in Fig. 2.2b) was superposed. The sustained response appeared to decay back to baseline between about 600 and 650 ms after the onset of the test portion (2,100-2,150 ms relative

to stimulus onset). Therefore the data were corrected for a baseline that was constant and equal to the average of the 200-ms period just before the onset of the test portion (1,300 - 1,500 ms) for times ≤ 2,100 ms, and then decreased linearly to zero between 2,100 and 2,150 ms.

The sources of the ITD change responses were modelled with a single equivalent current dipole in each hemisphere and a four-shell ellipsoidal volume conductor as head model using the Brain Electrical Source Analysis software (BESA, version 5.1.8; Megis, Gräfelfing, Germany). The dipole locations were fixed at the centroid of primary auditory area TE1.0 [Talairach co-ordinates: -47.5 -21.7 13.1 (left) and 50.5 -17.9 10.1 mm (right); Morosan et al., 2001], the Montreal Neurological Institute (MNI) coordinates of which were calculated with the SPM Anatomy toolbox (www.fz-juelich.de/ime/spm_anatomy_toolbox; Eickhoff et al., 2005) and converted to Talairach space using the non-linear transformation proposed by Brett et al. (2002; http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach), and the dipole orientations were fitted to the data.

Participants

A total of 10 participants (6 female and 4 male, age range: 19 - 34 years) took part in the experiment after having given written informed consent. All participants were right-handed (laterality indices equal to or greater than 57), as assessed through a modified version of the Edinburgh inventory (Oldfield 1971), and had no history of audiological or neurological disease. Participants were paid for their services at an hourly rate. The experimental procedures conformed with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and were approved by the Ethics

Committee of the University of Nottingham Medical School.

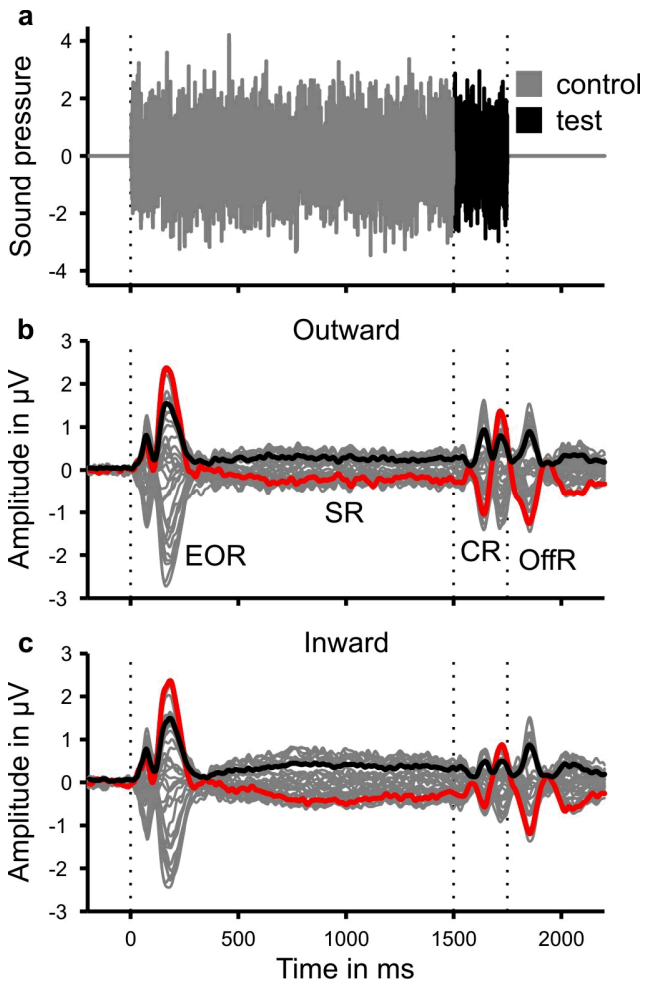

FIG. 2.2 Exemplary stimulus waveform (a), and grand-average responses (b,c), plotted as a function of time relative to stimulus onset. In panel a, the grey and black lines show the control and test portions of the stimulus, respectively. The responses in panels b and c were averaged across all three outward and inward conditions, respectively. The thin grey lines show the responses from all 33 electrodes. The black lines show the root mean square (rms) amplitude of the responses, and the red lines show the response from the vertex electrode (Cz). The vertical dotted lines mark the stimulus onset, the change in ITD between the control and test portions (at 1,500 ms), and the stimulus end (at 1,750

ms). EOR: energy onset response; SR: sustained response; CR: change response; OffR: offset response.

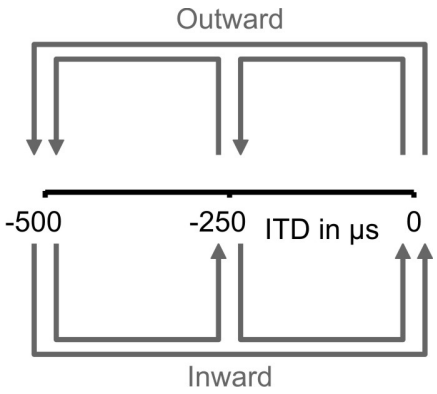

FIG. 2.3 Schematic representation of the six ITD change conditions used in the current study. The upper arrows show the three outward conditions , with ITDs changing from 0 to -250 or -500 µs, or from -250 to -500 µs. The lower arrows show the corresponding inward conditions (-250 to 0 µs, -500 to 0 µs and -500 to -250 µs).

2.C RESULTS

Comparison of average outward and inward responses

The average responses to both the outward and inward conditions (averaged over all three conditions within each category) exhibited a transient response to the onset of the control sound (labelled "energy onset response", or EOR, in Fig. 2.2b), a sustained response (SR), which remained roughly constant throughout the rest of the stimulus, a transient response to the ITD change at the transition from the control to the test portion (CR) and an off-response following the end of the stimulus (OffR). The ITD change response, which would be assumed to reflect ITD-specific processing in auditory cortex (Ungan et al., 1989, 2001; Jones et al., 1991), comprised three deflections: a small positive deflection (cP1 in Fig. 2.4a), followed by a large negative (cN1) and then another larger positive deflection (cP2). The negative deflection following the cP2 (labelled OffR in Fig. 2.4a) does not occur for shorter test sound durations (Hewson-Stoate et al., 2006; Krumbholz et al., 2007) and must thus be assumed to represent an off response to the end of the test sound. As the most prominent deflections in the ITD change response, the following analysis will focus mainly on the cN1 and cP2.

In accordance with the predictions of the hemispheric-channel model of ITD processing, both the cN1 and, to a lesser degree, also the cP2 deflection appeared to be larger for the outward than the inward conditions (compare panels A and B in Fig. 2.4). To test the statistical significance of this difference, we measured the peak-to-peak (pp) amplitude between the cN1 and cP2 deflections in the vertex channel (Cz) of the average outward and inward responses for each participant by calculating the difference between the average voltages within the 40-ms time windows centred around the root mean square

(rms) peaks of the deflections. A paired t-test of these pp amplitudes confirmed that the difference between the outward and inward conditions was significant [$t(9)= 5.978$, $p < 0.001$].

Hemispheric distribution of ITD change responses

The field maps of both the cN1 and cP2 deflections of the ITD change responses (averaged over the 40-ms windows around the rms peaks of the deflections) were consistent with a source in the region of auditory cortex (panels a, b, and c in Fig. 2.5). The field map of the cN1 deflection in the average outward response suggests that the outward response was biased towards the right hemisphere, contralateral to the perceived lateralisation of the test portion (Fig. 2.5a). The average cN1 deflection to the inward conditions was too small to result in a meaningful field map. However, the field map of the cP2 deflection in the average inward response suggests that, unlike the outward response, the inward response was biased towards the left hemisphere, *ipsilateral* to the perceived lateralisation of the control portion (Fig. 2.5c). As explained in the Sec. 2.A, this pattern of hemispheric lateralisation would be expected to be based on the hemispheric-channel model of ITD processing, because, in this model, only the ipsilateral channel would be expected to respond to an inward ITD change (see Fig. 2.1b). In contrast, the topographic model would predict the hemispheric distribution of the inward response to be similar to that of the outward response.

Inferring the hemispheric lateralisation of EEG responses from channel data can be misleading, so we used an equivalent current dipole source model to obtain a more reliable estimate of activation or source strength in each hemisphere (see Sec. 2.B). The dipole model was fitted to the average ITD change response for all conditions and

participants. The fitting was performed within the time window ranging from the start of the cN1 deflection to the end of the cP2 deflection, based on the respective rms minima (112 to 284 ms relative to the onset of the test portion). The cN1-cP2 pp amplitudes of the source waveforms for the left- and right-hemisphere dipoles (derived in the same way as for the channel data) confirmed that, while the outward response was strongly biased towards the right hemisphere (contralateral to the perceived lateralisation of the stimuli), the inward response showed a slight bias towards the left (ipsilateral) hemisphere (panels d and e in Fig. 2.5). This was corroborated by a repeated-measured analysis of variance (ANOVA) of the source amplitudes with factors ITD change direction (outward and inward) and hemisphere (left and right), which revealed a significant interaction between ITD change direction and hemisphere [$F(1,9) = 11.649$, $p = 0.008$]. The main effect of ITD change direction was also significant, as expected [$F(1,9) = 25.249$, $p = 0.001$; note different ordinate scales in panels d and e of Fig. 2.5]. The main effect of hemisphere was not significant [$F(1,9) = 3.407$, $p = 0.98$].

Response pattern to individual ITD changes

The results so far are more consistent with the hemispheric-channel than the topographic model, in that the average response to the inward ITD changes was considerably smaller and showed a more ipsilateral hemispheric distribution than the response to the outward changes. In this section, we show that the hemispheric-channel model is also consistent with the detailed pattern of results obtained for the individual ITD change conditions tested in this study.

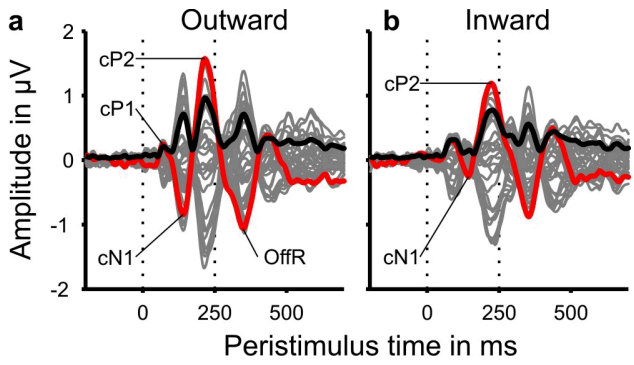

FIG. 2.4 Average ITD change responses to outward (a) and inward (b) conditions, plotted as a function of time relative to the ITD change. As in Fig. 2.2, the thin grey lines show the responses from all 33 electrodes. The black lines show the root mean square (rms) amplitude of the responses, and the red lines show the vertex responses (Cz). The vertical dotted lines mark the ITD change and the stimulus offset after 250 ms. cP1, cN1, and cP2: deflections of the ITD change response; OffR: offset response.

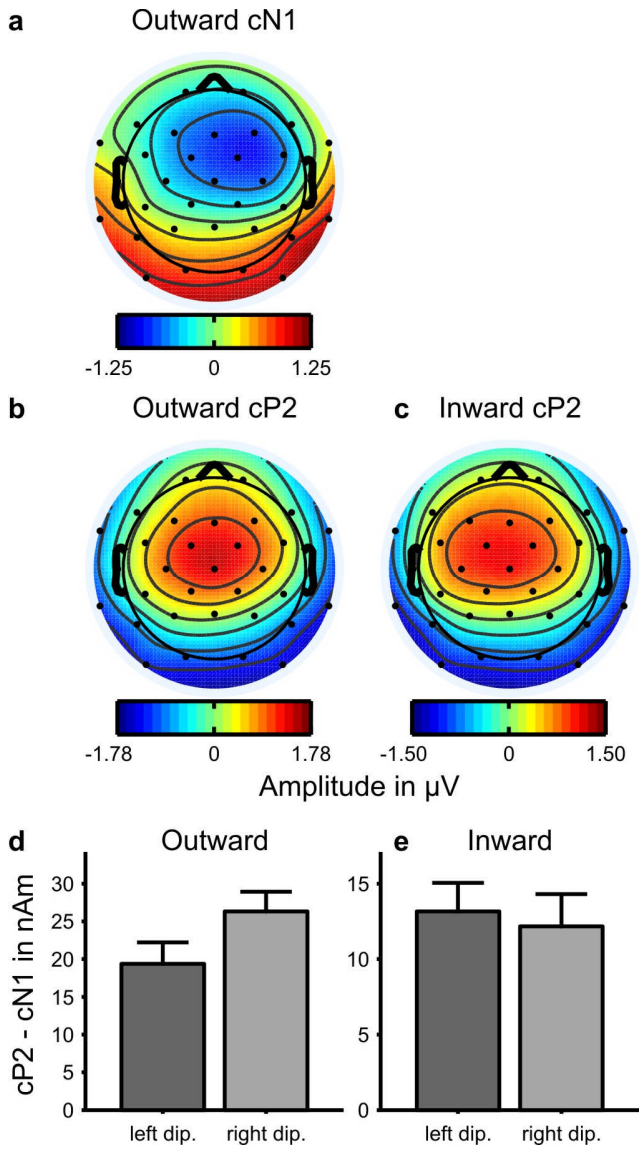

FIG. 2.5 Field maps (a-c) and source amplitudes (d,e) of average ITD change responses to outward and inward conditions shown in Fig. 2.4. The field maps were computed over the 40-ms windows around the root mean square (rms) peaks of the cN1 (a) and cP2 deflections (b,c) of the responses. The cN1 deflection of the average inward response was too small to yield a meaningful map. In panels d and e, the dark-grey and light-grey bars show the source amplitudes for the left- and right-hemisphere dipoles, respectively.

According to the hemispheric-channel model, the response amplitude to outward ITD changes should increase with the size of the change, because the stimulus would be moving further along the monotonically increasing ITD-response function of the relevant channel (the blue channel in Fig. 2.1b for the stimuli used in this study). In agreement with this prediction, the largest outward ITD change, from 0 to -500 μs, produced the largest response (black line in Fig. 2.6a and black bar on left side of Fig. 2.6c), and the other two outward ITD changes, from 0 to -250 and from -250 to -500 μs produced smaller and similar-sized responses (red and blue lines in Fig. 2.6a and bars on left side of Fig. 2.6c). For the inward ITD changes, the hemispheric-channel model would predict that most of the response should be generated by neural elements belonging to the channel ipsilateral to the perceived lateralisation of the control sound (the red channel in Fig. 2.1b for the stimuli used in this study), because only they could respond more strongly to the test than the control ITD. Therefore, the response amplitude to the inward ITD changes should be mainly determined by the ITD of the test portion. In particular, for the test ITD of -250 μs (in the -500 to -250 μs transition), the change response should be even smaller than for the zero test ITD (in the -500 to 0 and -250 to 0 μs transitions). The data were consistent with this prediction (see blue line in Fig. 2.6b and bar on right side of Fig. 2.6c). Furthermore, the responses to the two inward conditions with zero test ITD (-500 to 0 and -250 to 0 μs) should depend little on the control ITD and thus be of about the same amplitude. This prediction was also borne out by the data (compare red and black lines in Fig. 2.6b and bars on right side of Fig. 2.6c). These results were confirmed with a repeated-measures ANOVA of the vertex (Cz) cN1-cP2 pp amplitudes for the individual ITD change conditions with ITD change direction (outward or inward)

and condition (control and test ITDs of 0 and -250, 0 and -500, or -250 and -500 μs) as factors. Both main effects [change direction: $F(1,9) = 32.693$, $p < 0.001$; condition: $F(2,18) = 12.105$, $p < 0.001$] and the interaction [$F(2,18) = 15.342$, $p < 0.001$] were significant.

2.D DISCUSSION

The current results are consistent with the idea that, in humans, ITDs are coded by the activity levels in two broadly tuned hemispheric channels (hemispheric-channel model), rather than by the spatial distribution of activity across many finely tuned channels (topographic model). Using EEG and the continuous stimulation paradigm, this study showed that an ITD change towards the midline (inward change) produces a considerably smaller response than the reverse change away from the midline (outward change). Moreover, the inward response showed a more ipsilateral hemispheric distribution than the outward response, which was predominantly contralateral for the left-lateralised stimuli used in the current study. The hemispheric-channel model was also consistent with the detailed pattern of results for the individual ITD change conditions, with different starting and ending ITDs, tested in this study.

Using computer modelling, Harper and McAlpine (2004) showed that a topographic or intermediate code would optimise ITD discrimination performance at all but the lowest frequencies (< 250 Hz) in humans. However, discrimination performance is only one of several constraints to influence the evolution of the binaural system in humans, and hemispheric-channel coding may be superior to topographic coding for other functions, such as sound localisation or signal detection in noise. Moreover, an opponent-channel code of sound location would allow other stimulus features, such as

pitch or loudness, to be encoded by the same population of neurons without the need for interleaved, or nested, feature maps (Knudsen et al., 1987), and thus provide an efficient means for combining spatial with non-spatial information (Stecker et al., 2005).

Some perceptual models of interaural temporal processing assume an "inverted" topographic coding mechanism, whereby ITDs are represented by *minima*, rather than maxima, in neurons' ITD response functions (Breebaart et al., 2001; Durlach, 1972; Lindemann, 1986). Neurons that show a firing minimum at a consistent ITD value across frequencies are known to exist at all levels of the mammalian binaural system (e.g., Fitzpatrick and Kuwada, 2001; Fitzpatrick et al., 2002). The ITD response functions of these so-called "trough-type" neurons look like inverted versions of the schematic tuning curves shown in Fig. 2.1a. Unlike the original topographic model, the inverted version of the model would be able to account for the observed smaller size of the inward ITD change responses compared to the outward responses, if it is assumed that there are a greater number of neurons "tuned" (in terms of response minimum) to ITDs near the midline than to more lateral ITDs. However, the inverted topographic model would be inconsistent with the pattern of results found for the individual ITD change conditions tested, because the model would predict the size of the ITD change response to be exclusively determined by the size of the population tuned to the control ITD. This is because "trough-type" neurons would be expected to respond maximally to all ITDs outside of the "tuned" ITDs, and so the ITD change response would reflect the activity of neurons which were minimally active during the control portion. This inverted topographic model would predict that for the outward ITD changes, the response to the changes with a control portion of 0 μs (0 to -250 and 0 to -500 μs; see red and black bars

on left side of Fig 2.6c) should be similar, and both of these responses should be larger than the response to the ITD change with a control portion of -250 μs (blue bar). Similarly, for the inward conditions, the size of the ITD change response should be similar for the changes with a control portion of -500 (-500 to 0 and -500 to -250 μs; see black and blue bars on right side of Fig. 2.6c). None of these predictions are consistent with the data, and this means that the ITD change responses observed in the current study are unlikely to have arisen from ITD-sensitive neurons with trough-type response characteristics.

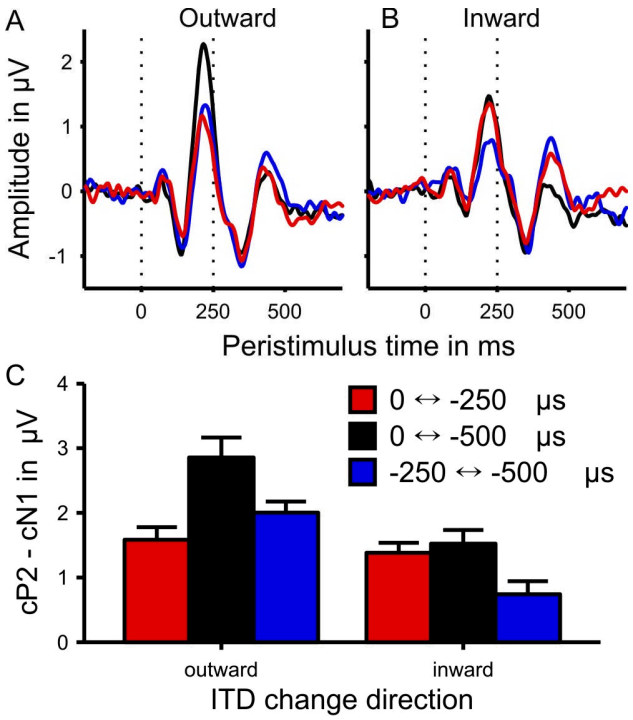

FIG. 2.6 a,b: Vertex (Cz) responses to individual ITD change conditions, plotted as a function of time relative to the ITD change. The responses to the outward conditions are shown in panel a and the inward responses are shown in panel b. Different line colours denote different conditions (red: 0 ↔ -250 µs; black: 0 ↔ -500 µs; blue: -250 ↔ -500 µs; see legend). Panel c shows the corresponding cN1-cP2 peak-to-peak amplitudes, using the same colour code as in the upper panels.

Physiological data indicate that the majority of ITD-sensitive neurons in each hemisphere are tuned to ITDs from the contralateral hemifield (e.g., McAlpine et al., 1996; Brand et al., 2002), suggesting that each hemisphere's response to ITDs is dominated by a single, contralateral ITD channel, and that ITDs are coded by the difference in activity level between these two hemispheric channels (McAlpine et al., 2001). However, Stecker et al. (2005) pointed out that, in an opponent-channel mechanism involving comparisons between the hemispheres, a unilateral lesion should produce localisation deficits throughout the entire acoustic field, when, actually, only the contralesional hemifield is affected in most cases (e.g., Malhotra et al., 2004). Thus, Stecker et al. proposed a four-channel mechanism, with one contralateral and one ipsilateral channel in each hemisphere. Based on the existing physiological data, the ipsilateral channel would be assumed to be much smaller (involve fewer neurons) than the contralateral channel in non-human mammals. In humans, neuroimaging and neuropsychological data suggest that the balance between contra- and ipsilateral channels may differ between the hemispheres. The neuroimaging data indicate that the right hemisphere responds about equally strongly to ITDs from both hemifields, whereas the left hemisphere predominantly responds to ITDs from the right hemifield (Krumbholz et al., 2005a). This suggests that contra- and ipsilateral channels may be more evenly balanced in the right than the left hemisphere. Electrophysiological data even suggest that the ipsilateral channel may be completely lacking in the left hemisphere, and that the left-hemispheric response to ITDs from the ipsilateral hemifield is relayed through callosal connections (Krumbholz et al., 2007). The idea that humans possess three ITD channels, two in the right and one in the left hemisphere, is consistent with neuropsychological

findings showing that right-hemisphere lesions generally lead to spatial processing deficits in both hemifields, whereas patients with left-hemisphere lesions usually perform normally in auditory spatial tasks (Clarke et al, 2000; Zatorre and Penhume, 2001). The idea that ITD processing in the left hemisphere is more reliant on callosal input than in the right hemisphere is consistent with findings of sound lateralisation deficits in patients with callosotomy, indicating a significant leftward bias in the lateralisation judgments in these patients (Hausmann et al., 2005).

Chapter 3. A new paradigm for measuring feature-specific auditory cortical responses with rapid event-related fMRI[3]

3.A INTRODUCTION

The response to an ITD change in an otherwise continuous sound, as measured in **Chapter 2**, would be assumed to reflect the response of those neural elements that are specifically involved in the processing of interaural temporal information. However, the response could also be due to an unspecific change detection mechanism of the type suggested to underlie the auditory oddball or mismatch response (for review, see Näätänen and Winkler, 1999), or to a generic "edge" detection process, related to the perception of auditory objects, as suggested by Chait and co-workers (2008; see also Chait et al., 2007). In order to investigate this question, the current study used functional magnetic resonance imaging (fMRI) to measure the topography of the ITD change response measured in **Chapter 2**.

FMRI studies have greatly increased our understanding of the functional organization of the human auditory cortex (for review, see Hall et al., 2003). Most of these studies have used epoch-related or "blocked" designs to investigate regional effects associated with the processing of specific sound features, such as pitch or spatial location (e.g., Baumgart et al., 1999; Patterson et al., 2002; Warren et al., 2002; Warren and Griffiths, 2003; Penagos et al., 2004; Krumbholz et al., 2005a,b; Hall and Plack, 2009). In such designs, the response to a test stimulus is compared with the response to an appropriately matched control stimulus, where the test stimulus possesses the relevant feature (e.g., pitch), while the control stimulus does not. The control response is

[3] Based on Magezi and Krumbholz (2009a)

subtracted from the test response to isolate activity associated with the processing of the test feature.

Neural responses in auditory cortex can be broadly classified into transient and sustained components (e.g., Recanzone, 2000; Eggermont, 2002). Transient and sustained response dynamics are reflected not only in the spiking activity, but also in the slow extracellular potentials relating to synaptic and other sub-threshold activity (Logothetis and Wandell, 2004), the far-field components of which can be measured through electro- or magneto-encephalography (EEG/MEG; Scherg and Picton, 1991), and in the resulting hemodynamic response, which underlies the blood oxygen level-dependent (BOLD) effect in fMRI (Logothetis et al., 2001; Seifritz et al., 2002). In epoch-related fMRI designs, the test and control stimuli are presented for prolonged periods of time (epochs), lasting several seconds to a few tens of seconds. Thus, the resulting activations mainly reflect the sustained components of the stimulus-related responses (Harms and Melcher, 2002; 2003; Harms et al., 2005). This is particularly true when using a sparse image acquisition protocol as is often the case in auditory fMRI (Hall et al., 1999). In contrast, electro- or magneto-encephalographic (EEG/MEG) studies of auditory processing usually focus on the transient response components elicited at stimulus onset. In order to compare results from fMRI and EEG or MEG, and exploit the complementary strengths of these different methodologies with respect to spatial and temporal resolution, it would be desirable to probe functionally equivalent aspects of the stimulus-related responses by using the same or similar experimental paradigms with both kinds of methodologies. Applying the subtraction approach from epoch-related fMRI designs to EEG or MEG would be problematic, because subtraction can severely degrade the signal-to-noise ratio

of the difference response (Fig. 3.1a) and may create artifactual deflections if the control and test responses differ in latency (Fig. 3.1b). As shown in **Chapter 2**, in EEG and MEG, activity associated with feature-specific processing can be isolated without the need for subtraction by preceding the test sound directly with the control sound and measuring the response to the transition between the two sounds (e.g., Halliday and Callaway, 1978; Jones et al., 1991; Martin and Boothroyd, 1999, 2000; Krumbholz et al., 2003; Chait et al. 2005; Krumbholz et al., 2007). In this paradigm (referred to as the continuous stimulation paradigm, or CSP; see Sec. 2.A), the test sound is usually presented for only a brief period of time (a few hundred milliseconds), whereas the control sound has a longer duration (a few seconds) to allow the transient response to the onset of the control sound to subside before the transition to the test sound. The response to the transition from the control to the test sound would then be assumed to reflect processing related to the perceptual change (e.g., feature onset) that occurs at the transition.

The current study applies the CSP to auditory fMRI. The aim was to test the above assumption by measuring the topography of the transition response to different feature onsets and comparing the results to those from epoch-related studies. In visual fMRI, the CSP has been applied to investigate selectivity for orientation (Tootell et al., 1998) and motion direction (Tootell et al., 1995) in the human visual cortex. The CSP is also related to fMRI paradigms based on adaptation, that is, the reduction in brain activity when stimuli are repeated (for review, see Grill-Spector et al., 2006). As adaptation is stimulus specific, a change in stimulus tends to produce an enhanced response compared to a repeated stimulus. Like the transition response in the CSP, this release from

adaptation is thought to reflect processing specifically related to the stimulus change. So far, adaptation-based fMRI paradigms have mainly been used in the visual domain. Another related paradigm is the auditory oddball paradigm, which measures the mismatch response to infrequent deviant stimuli in an otherwise repetitive sequence of standard stimuli (Näätänen et al., 1978). This paradigm can be used to study auditory processing with both fMRI and EEG or MEG (Opitz et al., 1999a,b, 2002; Doeller et al., 2003; Molholm et al., 2005; Schönwiesner et al., 2007b). Unlike adaptation-based or oddball paradigms, the CSP avoids the need for subtraction, and is thus likely to yield a better signal-to-noise ratio. The CSP would also seem an excellent candidate paradigm for simultaneous EEG-fMRI measurements (for review, see Herrmann and Debener, 2008). In the current study, the CSP was used to measure transient fMRI responses associated with the processing of pitch and sound motion in the human auditory cortex. Previous epoch-related fMRI studies indicate that the sustained activity associated with the processing of these features involves different areas in auditory cortex. The aim of the current study was to test whether the same would apply to the respective transient responses.

3.B METHODS

Stimuli

The stimuli consisted of alternating sequences of control and test sounds, starting with a control sound, and were presented continuously over time (Fig. 3.2). As in previous electrophysiological experiments (see e.g., Sec. 2.B; Krumbholz et al., 2007), the control sounds had a relatively long duration of 2010 ms and the test sounds had a much shorter duration of 300 ms; thus, the stimulus onset asynchrony (SOA) between

successive test sounds was 2310 ms. All sounds were based on random noise and had the same gross spectral and temporal characteristics. The control sounds consisted of random Gaussian noise, which was multiplied with a lowpass-noise envelope (100-Hz cutoff, 4^{th}-order Butterworth) to make the envelope spectrum of the noise more similar to that of natural sounds like speech and music (Singh and Theunissen, 1998), and lowpass-filtered at 1 kHz (4^{th}-order Butterworth) to minimize spectral overlap with, and thus perceptual masking by, the scanner noise (Gaab et al., 2007b). There were four different test sounds, which were presented in a predefined pseudorandom order (see Experimental protocol in this section). In the first condition, referred to as the "null" condition (labelled "N" in Fig. 3.2b), the test sound was simply a continuation of the preceding control sound. In the second condition, referred to as the "pitch" condition ("P" in Fig. 3.2b), the test sound was an iterated rippled noise (IRN). IRNs, which are often referred to as regular-interval sounds in the neuroimaging literature (e.g., Patterson et al., 2002), are created by delaying a copy of a random noise, adding it back to the original and iterating the process several times. The resulting sound elicits a buzzy pitch corresponding to the reciprocal of the delay (Yost, 1996). In the current experiment, the IRN test sounds were produced with a delay of 4 ms, corresponding to a pitch of 250 Hz, and 8 iterations of the delay-and-add process. They were lowpass-filtered in the same way as the control sounds (1– kHz cutoff). In order to avoid audible transients at the transitions from the control to the pitch test sounds, the IRN test sounds were gated on and off with 5-ms cosine-squared gates and cross-faded with the surrounding control sounds so that the energy envelope of the composite sound remained flat. In the third and fourth conditions, referred to as the "motion left" and "motion right" conditions ("ML" and "MR" in Fig. 3.2b), the test

sounds were a continuation of the preceding control sound as in the null condition. In this case, however, the interaural time difference (ITD) was varied over time to create the sensation of motion. In order to avoid ITD discontinuities at the transitions with the surrounding control sounds (which had zero ITD), the ITD was increased linearly from 0 to 1000 μs (leading at the left or right ear, respectively) over the first half of the test sounds (150 ms) and then decreased back to 0 μs again over the second half (see bold, black line and right ordinate in Fig. 3.2d,e). Given that the ITD was applied after multiplication with the lowpass-noise envelope, it was represented both in the temporal fine structure and the envelope of the waveform to maximize its perceptual salience (Nuetzel and Hafter, 1976). Stepwise discontinuities in the ITD relating to the sampling period were avoided by linearly interpolating between waveform samples.

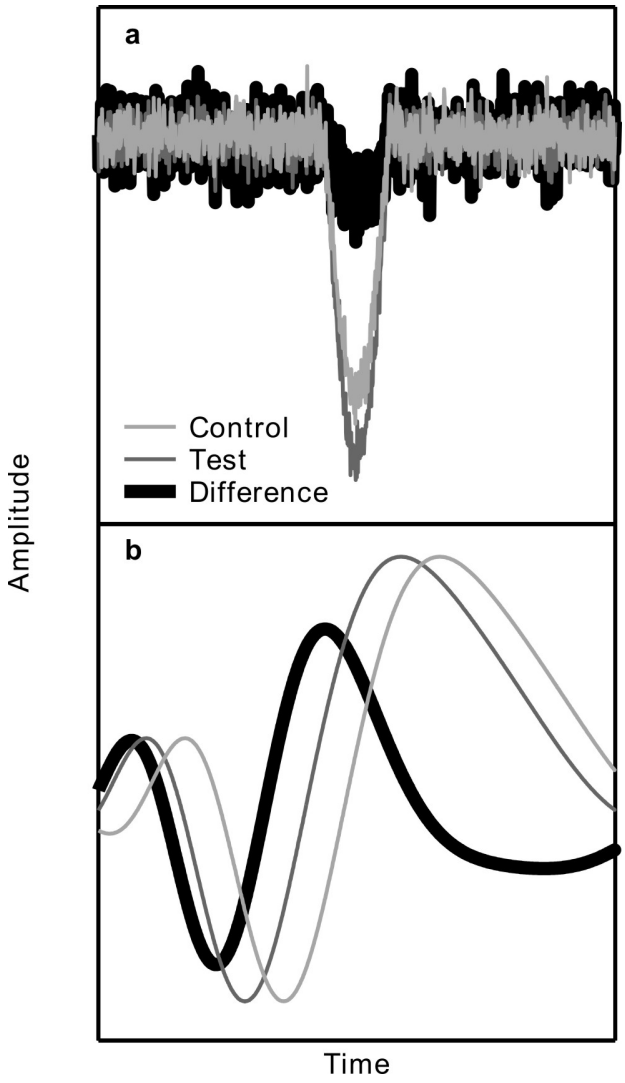

FIG. 3.1 Simulated electrophysiological (EEG or MEG) responses illustrating the pitfalls of applying the subtraction approach to EEG or MEG data. The thin lines show simulated responses to a control (light grey) and a test (dark grey) stimulus, plotted in arbitrary units as a function of time. The bold black lines show the difference between the test and control responses (test - control). (a) Noise in the test and control responses, simulated by adding random Gaussian noise, usually results in a much degraded signal-to-noise ratio

of the difference response. (b) A latency difference between the test and control responses leads to artifactual deflections in the difference response.

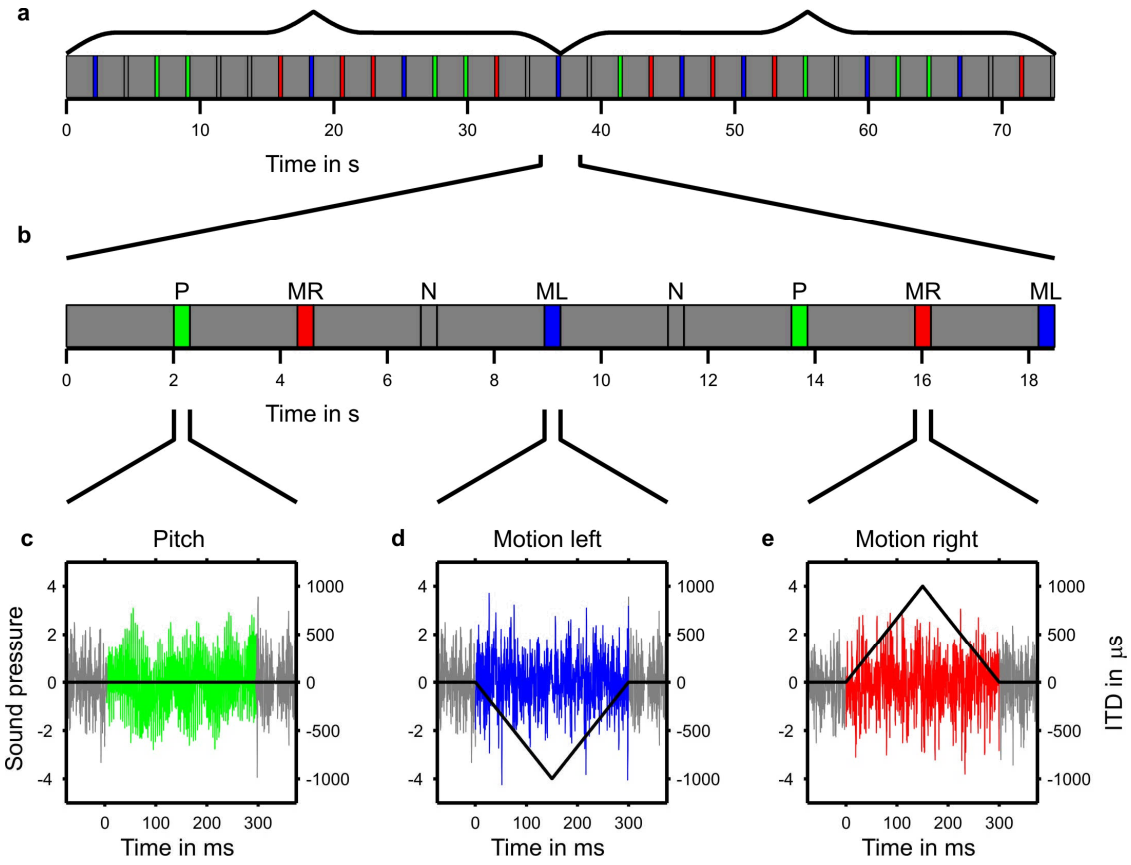

FIG. 3.2 Experimental stimuli. (a,b) Schematic representation of a sequence of control (grey background) and test sounds (coloured insets). The null events ("N" in panel b) are represented by grey, the pitch event ("P") by green and the motion-left ("ML") and -right ("MR") events by blue and red insets. Four presentations of each test sound were randomly permuted within successive blocks (curly brackets in a). (c-e) Exemplary waveforms (thin lines and left ordinate) of the pitch (c), motion-left (d) and motion-right (e) events, plotted as a function of time relative to the onset of the test portion (shown in

colour). The bold black lines show the instantaneous ITD of the stimuli in μs (right-ordinate).

The stimuli were generated digitally and digital-to-analogue converted with a 24-bit amplitude resolution and a 12-kHz sampling rate using TDT System 3 (Tucker Davies Technologies, Alachua, FL, USA) and MATLAB® (The Mathworks, Natick, MA, USA). They were delivered via the magnetic resonance-compatible sound system developed by the MRC Institute of Hearing Research (Palmer et al., 1998 ; http://www.ihr.mrc.ac.uk/research/technical/soundsystem/index.php), which consists of high-quality electrostatic transducers (HE60, Sennheiser, Wedemark, Germany) fitted into professional ear defenders (Bilsom 2452) to provide passive shielding from the scanner noise. The control and test sounds were presented at the same overall level of about 75 dB SPL and were well audible over and above the scanner noise. The noise for the control and test sounds was continually created afresh by means of double buffering.

Experimental protocol

The experiment consisted of two runs, each of which lasted about 22.5 min and contained 144 presentations of each of the four test sounds (null, pitch, motion left and motion right). The SOA between successive test sounds (2310 ms) was much shorter than the duration of the hemodynamic response that each test sound would be expected to evoke (approximately 25 s; Blamire et al., 1992). In order to be able to deconvolve the overlapping hemodynamic responses to the test sounds and recover the response function to each event, the different test sounds had to be presented in a randomized order (Dale and Buckner, 1997). The efficiency of such rapid event-related designs has been shown to depend not only on the SOA between successive events and the image acquisition rate, but also on the exact order of presentation of the different event types (Friston et al., 1999). In the current experiment, the different event types (null, pitch, motion left and

motion right) were presented in pseudorandom sequences selected to optimize design efficiency. A new sequence was generated for each participant and each run. Sequences were created by randomly permuting 16 test events (four presentations of each of the four test sounds) within successive 36.96-s blocks (16×2310 ms = 36.96 s; see curly brackets in Fig. 3.2a). This was to prevent excessive over- or under-representation of particular events within particular periods during the experiment, and the associated reductions in efficiency they cause (Friston et al., 1999). For each sequence, efficiency scores for each event type (pitch, motion-left and –right) were calculated as described in Friston et al. (1999). Only sequences with above-average efficiency scores (distribution based on 1000 sequences) were selected for the experiment.

fMRI scanning was continuous with an image repetition time of TR = 2 s. The start of the stimulus presentation in each run was synchronized with the beginning of the first image by means of a hardware trigger. Participants watched a self-chosen silent movie through fibre-optic goggles (SV – 7021, Avotec, Stuart, FL, USA) to maintain wakefulness throughout the experiment.

Data acquisition

Blood oxygen level-dependent (BOLD) contrast images were acquired with a Philips 3-T Intera whole-body scanner equipped with an 8-channel sense head coil (Philips, Eindhoven, The Netherlands) and gradient echo-planar imaging (EPI; data matrix: 200×200, TR = 2 s, TE = 50 ms, flip angle = 90°). The functional images consisted of 18 descending slices with an in-plane resolution of 2.5×2.5 mm^2, a slice thickness of 2.5 mm and no inter-slice gap. They were oriented so that the slices would be approximately parallel to the Sylvian fissure, and the middle of the volume positioned

at the base of Heschl's gyrus. The positioning of the functional slices was performed with a high-resolution structural scan, which was acquired at the beginning of the experiment for each participant (MPRAGE; 1 mm^3 isotropic resolution, data matrix: 256 x 256, TR = 8.1 ms, TE = 3.7 ms, min. TI = 867 ms, shot interval = 3 seconds, flip angle = 8°, SENSE factor 3). A total of 676 functional images were acquired in each run. Each run was preceded by 4 dummy images to allow for magnetic saturation.

To aid co-registration of the functional and structural images for data analysis, a whole-head EPI image was also acquired using the same imaging parameters as for the functional images, apart from the echo time (TE), which was 30 ms instead of 50 ms. The whole-head EPI image consisted of 60 slices, which had the same orientation and middle position as in the functional images.

Data analysis

Functional and structural images were analyzed using SPM2 (http://www.fil.ion.ucl.ac.uk/spm). The functional images were slice-time corrected, using the 9^{th} slice as reference slice, realigned to the first image of the first run to correct for head motion, and averaged to create an average functional image for co-registration with the structural image. The whole-head EPI image was co-registered with the structural image and then the average functional image was co-registered with the whole-head EPI image. The co-registration parameters for the average functional image were then applied to the individual functional images. The structural image was normalized to a symmetrical version of the Montreal Neurological Institute (MNI) T1 standard template (ICBM152) in SPM. A symmetrical template was used, because one analysis involved comparing contrast images across hemispheres by contrasting the original images with

left-right flipped versions of the same images. Unless a symmetrical template is used, such comparisons would be sensitive to any inter-hemispheric differences in the normalization process (e.g., Watkins et al., 2001; Jäncke et al., 2002; Krumbholz et al., 2009b). The symmetrical template was created by averaging the original with a flipped version of the standard template. The normalization parameters were applied to the functional images and the functional images were spatially smoothed using a Gaussian kernel with 10-mm full width at half maximum.

The data of each participant were modelled with a general linear model (Friston et al., 1995), which included regressors for all test events apart from the null events (i.e., pitch, motion left and motion right). The BOLD responses to these events were modelled by convolving a delta or "stick" function at the onset of each test sound with a canonical hemodynamic response function (Friston et al., 1998). As the null events were modelled implicitly, the transition responses to the pitch and motion onsets were obtained without any explicit subtraction. The data were high-pass filtered with a cut-off period of 128 s to remove low-frequency drifts, and serial correlations were accounted for.

Based on previous neuroimaging (e.g., Krumbholz et al., 2005a), physiological (e.g., Fitzpatrick and Kuwada, 2001) and lesion data (e.g., Lomber et al., 2007), we expected the responses to the motion-left and motion-right conditions to be lateralized to the hemisphere contralateral to the hemispace to which the moving sounds were lateralized. This was tested by flipping the individual contrast images for the motion-left and motion-right regressors by 180° about the mid-sagittal plane, and then subtracting the flipped contrast images from the original images.

For group analysis, contrast images for each participant were submitted to a voxel-wise one-sample *t*-test (one-tailed) and the resulting random-effects *t*-maps were thresholded at a voxel threshold of $t = 3.93$ ($p \leq 0.001$, uncorrected). In order to combine the motion-left and motion-right contrasts, the OR conjunction of the two contrasts was computed using the same voxel threshold of $t = 3.93$ ($p \leq 0.001$, uncorrected). Whereas the more commonly used AND conjunction of two contrasts is defined by the minimum of their *t*-maps (Nichols et al., 2005), the OR conjunction is defined by their maximum (see, e.g., Krumbholz et al., 2009b). The OR conjunction, rather than a simple average, was used to combine the contrasts because the motion-left and motion-right activations were largely non-overlapping. Averaging the contrasts would thus have unduly reduced the signal-to-noise ratio in the combined contrast.

For visualization, the statistical maps were projected onto oblique axial slices oriented parallel to the Sylvian fissure (30º pitch angle) of the MNI single-subject template (Colin27). The position of the slice was chosen as the average of the coordinates of the most significant voxels for the pitch and motion-related activations with respect to the normal to the plane of the slice (see Table 3.1).

Meta-analysis

In order to compare the transient event-related fMRI responses measured in the current experiment with the responses obtained from previous epoch-related fMRI studies, a meta-analysis was conducted of previous epoch-related fMRI studies of pitch and motion processing. As activation to spatial attributes of stationary sounds has been shown to comprise the same or similar areas as activation to sound motion (see, e.g., Krumbholz et al., 2005a, and references therein), studies on the processing of stationary

spatial cues were also included in the meta-analysis. The current results were also compared with results from a study by Gutschalk et al. (2004), which measured the MEG response to the onset of pitch in an otherwise continuous sound using a similar paradigm as in the current fMRI study. Gutschalk et al. used regular and irregular click-trains as stimuli, rather than IRNs and noises. Finally, an epoch-related fMRI study of the auditory oddball response to pitch and location deviants (Deouell et al., 2007) was also included in the meta-analysis.

Previous fMRI studies of pitch processing have used a variety of different stimuli and contrasts. Several studies have used IRNs like the current study (Patterson et al., 2002; Warren and Griffiths, 2003; Barrett and Hall, 2006); others have used spectrally resolved and/or unresolved complex tones (Penagos et al., 2004; Deouell et al., 2007). In some studies, the pitch response was contrasted with the response to an atonal sound, such as noise (Patterson et al., 2002), or a sound with a different pitch salience than the test sound (Penagos et al., 2004). Some studies contrasted the response to changing pitch with the response to a fixed pitch (Warren and Griffiths, 2003; Deouell et al., 2007).

Previous fMRI studies of spatial processing contrasted the response to different fixed or changing locations (Warren and Griffiths, 2003; Krumbholz et al., 2005a; Deouell et al., 2007), or to moving sounds (Warren et al., 2002; Krumbholz et al., 2005a,b), with the response to a fixed reference location. The perceived spatial location of the stimuli was manipulated using virtual acoustic space techniques (Wightman and Kistler, 1989) in all of these studies except for the studies by Krumbholz et al. (2005a,b), which used ITDs only, to exclude spectral cues. The stimuli were broadband as in the current study.

The meta-analysis was based on the reported coordinates of the most significant voxels of the pitch or spatial processing-related activations within each hemisphere in all but two of the studies considered here: Penagos et al. (2004) reported the coordinates of the "centre of mass" of activation clusters, rather than most significant voxels, and Gutschalk et al. (2004) reported the locations of equivalent dipole sources fitted to their MEG responses. Where the reported locations were based on the template of Talairach and Tournoux (1988), the coordinates were converted into MNI space using a non-linear transformation (http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach; Brett et al., 2002). The results of the meta-analysis were displayed on a similar oblique axial slice of the MNI Colin brain as the statistical maps of the current study (30º pitch angle). In order to display all data points on the same slice, the coordinates were projected along the normal to the plane of the slice. The location of the slice was chosen as the average of the coordinates with respect to the plane normal of all data points.

Participants

Thirteen volunteers (8 male and 5 female, aged between 21 and 50 years) with no history of audiological, psychiatric or neurological disease took part in the study after having given written informed consent. According to the Edinburgh inventory (Oldfield, 1971), nine volunteers were right handed [with laterality indices (LIs) equal to or greater than 38], one was left-handed (LI = -55), and three were ambidextrous (with LIs between -4 and 15). Participants who were not authors of the corresponding manuscript (Magezi and Krumbholz, 2009a) were paid for their services at an hourly rate. The experimental procedures were approved by the ethics committee of the University of Nottingham Medical School.

3.C RESULTS AND DISCUSSION

The current study used a rapid event-related design to measure the responses to pitch and motion onsets in an otherwise continuous sound. Rapid event-related designs offer greater efficiency than slow event-related designs involving sparse imaging, because more stimuli can be presented, and more images acquired, in a given period of time (Nebel et al., 2005). However, the continuous imaging in rapid designs may reduce stimulus-related responses in auditory cortex due to activation produced by the scanner noise (Shah et al., 1999, 2000; Novotski et al., 2001, 2006; Gaab et al., 2007a,b). Nevertheless, both the pitch and motion test events produced robust activations (Fig. 3.3a) with *t*-values comparable to those observed in previous studies of pitch and motion processing that have used epoch-related designs and sparse imaging (Table 3.1; e.g., Patterson et al., 2002; Warren and Griffiths, 2003; Krumbholz et al., 2005a,b). This suggests that transient event-related auditory responses may be less affected by scanner-noise related activity than sustained responses. This would seem plausible, because the scanner noise from continuous imaging would be expected to produce sustained activity, and previous results suggest that transient and sustained activity in auditory cortex may be spatially dissociable (Seifritz et al., 2002).

The activation associated with the pitch events (green highlight in Fig. 3.3a) was stronger and more widespread than the motion-related activation (magenta highlight; see also Table 3.1). It comprised the planum polare (PP), which is the part of the supratemporal plane (STP) anterior to the primary auditory cortex on Heschl's gyrus (HG), the antero-lateral and central parts of HG, as well as the antero-lateral part of the planum temporale (PT), which is located posterior to HG. In contrast, the motion-related

activation was mainly limited to the antero-central part of the PT and the central part of Heschl's sulcus. Comparing the activation patterns for the pitch and motion regressors with 50%-probability maps of the three cyto-architectonic subdivisions of primary auditory cortex in humans, TE1.0, TE1.1 and TE1.2 (Morosan et al., 2001; see inset in Fig. 3.3a) using the SPM Anatomy toolbox (www.fz-juelich.de/ime/SPM_anatomy_toolbox; Eickhoff et al., 2005) revealed that the pitch-related activation almost completely overlapped all three subdivisions in the left hemisphere (TE1.0: 100% overlap; TE1.1: 98%; TE1.2: 99%), and the central and anterior areas TE1.0 (99%) and TE1.2 (99%) in the right hemisphere. In the right hemisphere, the most posterior of the three areas, TE1.1, was only partially overlapped (33%), indicating that the pitch-related activation extended more posteriorly in the left than in the right hemisphere. This may relate to the fact that HG is located more anteriorly in the right than the left hemisphere (Leonard et al., 1998); this difference may not be adequately corrected by the normalisation process. The motion-related activation also extended more posteriorly in the left than in the right hemisphere. The motion-related activation had a fairly substantial overlap with the most posterior area TE 1.1 (54%) and marginally overlapped the central area TE 1.0 (12%) in the left hemisphere, but only overlapped the central area TE 1.0 (35%) in the right hemisphere. Importantly, there was no overlap with the most anterior area, TE1.2, in either hemisphere.

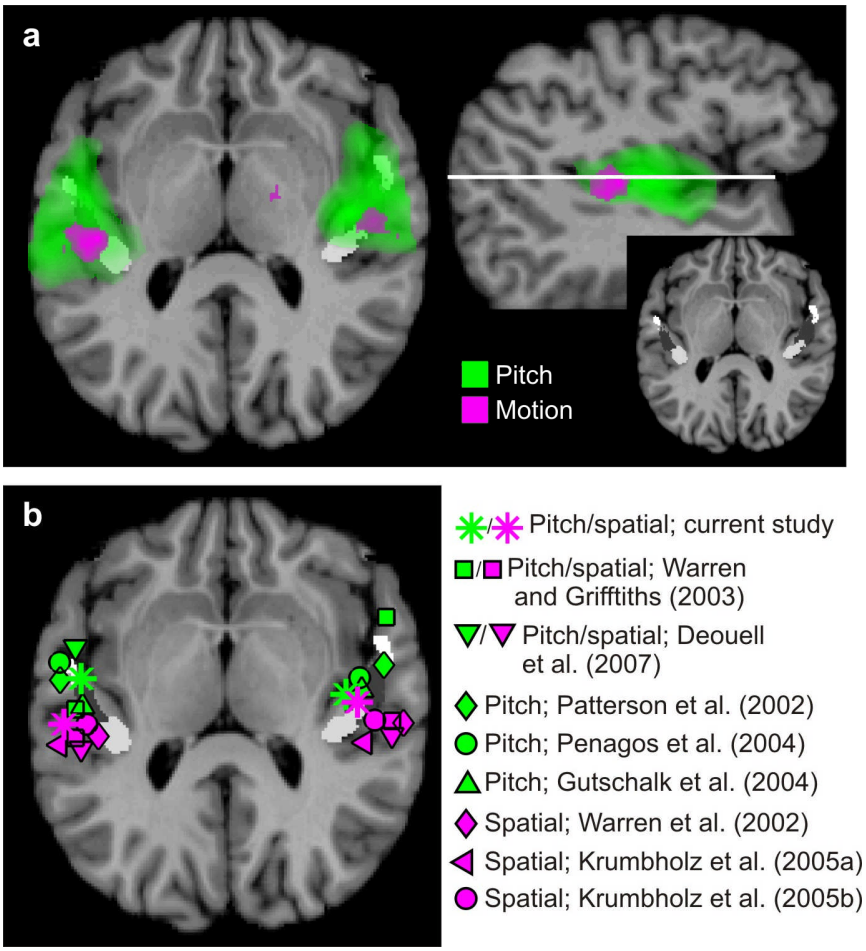

FIG. 3.3 Event-related activation to pitch and motion onsets and comparison with previous epoch-related studies. (a) Activation for the pitch regressor (green), and the OR conjunction of the motion-left and motion-right regressors (magenta). Data are shown on an oblique axial (parallel to the Sylvian fissure) and a sagittal (x = 46 mm) slice of the MNI single-subject template with 50%-probability maps of the three cyto-architectonic subdivisions of primary auditory cortex (TE1.0: dark grey, TE1.1: light grey, TE1.2: white; Morosan et al., 2001). (b) Meta-analysis of epoch-related fMRI studies on pitch (green) and auditory spatial processing (magenta), projected onto an oblique axial slice

[see section on meta-analysis in Methods (Sec. 3.B)]. The results from the current study are shown by stars (most significant voxels; see Table 3.1).

Table 3.1

MNI coordinates in mm and *t*-values of most significant voxels, as well as number of voxels in cluster (*k*) for event-related activations and hemispheric comparisons (see Sec 3.B)

Contrast	Brain region	Coordinates x, y, z	*t*-value	k
Pitch	Right HG	44, 16, 6	17.03	2909
	Left HG	-48, -12, 2	14.04	3255
Motion left	Left HS and PT	-44, -24, 8	5.21	97
	Right HS and PT	48, -18, 8	4.51	80
Motion right	Left HS and PT	-54, -28, 6	6.18	218
Motion left - flipped	Right HS and PT	60, -14, 12	6.38	104
Motion right - flipped	Left HS and PT	-40, -30, 10	5.31	40

HG: Heschl's gyrus; HS: Heschl's sulcus; PT: planum temporale

In the current study, the pitch-related activation was found to be much more widespread than and largely overlapping the motion-related activation. In contrast, most of the previous epoch-related studies have found pitch-related activation to be mainly restricted to the antero-lateral part of HG and to show little overlap with activation associated with spatial processing (e.g., Patterson et al., 2002; Warren and Griffiths, 2003; Barrett and Hall, 2006). The motion-related activation in the current study was very spatially specific, suggesting that this discrepancy is unlikely to reflect a general difference in specificity between the current event-related and previous epoch-related designs. It is important to note that the current study did not statistically analyse the pitch- and motion-related activations purely in terms of spatial distribution, that is, independent of activation strength.The difference may in part be due to improvements in the sensitivity of fMRI scanning techniques over recent years (Frahm et al., 2004). However, the main reason for the difference is probably related to differences in the stimuli used. Hall and Plack (2009) have shown that the topography of pitch-related activation is not independent of the nature of the stimulus used to elicit the pitch. Previous studies that have used IRN stimuli (e.g. Patterson et al., 2002; see meta-analysis in Sec. 3.B) have filtered the stimuli so they would mainly contain spectral components (or “ripples”; the harmonic components of IRN stimuli have finite bandwidth) that are unresolved by the cochlear filters. In contrast, the IRN stimulus used in the current study contained only resolved components and encompassed the main part of the perceptual dominance region for pitch (Ritsma, 1967; Dai, 2000). Resolved or dominant components are known to produce a much more salient and musical pitch than unresolved components (Houtsma and Smurzynski, 1990; Krumbholz et al., 2000; Pressnitzer et al.,

2001; Bernstein and Oxenham, 2003), and would thus be expected to produce stronger and more widespread activation.

Previous epoch-related fMRI studies (Warren and Griffiths, 2003; Barrett and Hall, 2006; for review, see Arnott et al., 2004) support the hypothesis that the primate auditory cortex is divided into two processing streams, an anterior or "what" stream, which is specialized in processing object-related information, and a posterior or "where" stream, which is assumed to be mainly concerned with spatial processing (Kaas and Hackett, 2000; Rauschecker and Tian, 2000; Tian et al., 2001). In order to test whether the transient event-related responses measured in the current study are consistent with this hypothesis, the most significant voxels in the pitch- and motion-related activations from the current study were compared with previous results from epoch-related fMRI studies of pitch and spatial processing (see section on meta-analysis in Sec. 3.B). Figure 3.3b shows that the results are remarkably consistent; in both the current and previous data, the peak in the pitch-related activation (green symbols in Fig. 3.3b) occurs anterior to the peak in the activation associated with spatial processing (magenta symbols). Interestingly, the peaks of the pitch-related activations seem to have a larger spatial spread across studies, reaching all the way from PP to the anterior PT, than the peaks of the activations related to spatial processing, which all seem to be confined to the antero-central PT. This is consistent with the idea that pitch processing can activate different areas within a larger region depending on the exact nature of the stimulus used (Hall and Plack, 2009). The close agreement between the current results and the results from previous epoch-related fMRI studies suggests that the transient and sustained components of the responses to pitch and motion or spatial processing have the same or similar

topographies. This finding is consistent with previous MEG results, which suggest that transient and sustained pitch-specific MEG responses arise from a similar area in auditory cortex (Gutschalk et al., 2004). Comparison with the current data indicates that the equivalent dipoles of the MEG responses co-localize with the most significant voxels in the fMRI activations. The current data are also in good agreement with previous fMRI data on the oddball response to pitch and location changes (Deouell et al., 2007).

In the motion conditions used in the current study, the test sounds were perceived as moving within either the left (motion left, see Fig. 3.2d) or the right hemispace (motion right, Fig. 3.2e), based on whether the changing ITD of the test sounds favoured the left or right ear (see Sec. 3.B). Based on previous imaging (e.g. Krumbholz et al., 2005a), physiological (e.g., Fitzpatrick and Kuwada, 2001) and lesion data (e.g., Lomber et al., 2007), the motion-related responses were expected to be lateralised to the hemisphere contralateral to the hemispace to which the moving sounds were lateralised. In order to test for contralateral asymmetry in the motion-related responses, the contrast images for the motion-left and motion-right conditions were compared with the respective left-right flipped versions of these contrasts (see Sec. 3.B). The comparison between a contrast and a flipped version of the same contrast can reveal whether the activation at a given voxel in one hemisphere is significantly greater than the activation at the corresponding voxel in the other hemisphere. This analysis showed that both the motion-left and the motion-right conditions produced a predominantly contralateral response (blue and red highlight in Fig. 3.4). In the case of the motion-left condition, 113 voxels showed a significantly stronger activation in the contralateral (right) than in the ipsilateral (left) hemisphere, and only 7 voxels were more strongly activated on the

ipsilateral side. Similarly, the motion-right condition produced a significantly stronger contralateral (left) than ipsilateral activation in 43 voxels, with no voxels being more strongly activated in the ipsilateral hemisphere. Previous studies in humans have often found a difference in the degree of the contralaterality of response between left- and right-lateralized sounds, in that the response to right-lateralized sounds tended to be more bilateral than the response to left-lateralized sounds (Deouell et al., 1998; Kaiser et al., 2000; Krumbholz et al., 2005a; Schönwiesner et al., 2007a; Krumbholz et al.,2007; Hine and Debener, 2007). While the current motion responses exhibited contralateral asymmetry, there was no apparent difference in the degree of contralaterality between the motion-left and motion-right conditions. The absence of such difference in the degree of contralaterality in the current study could reflect variability between different participant groups. This seems probable since hemispheric functional lateralization can differ greatly between individuals (Zatorre and Penhume, 2001). Alternatively, it could be due to stimulus-related or attentional factors (see, e.g., Schönwiesner et al., 2007a). Finally, it is also possible that differences in the degree of contralaterality of response between the motion-left and motion-right conditions were only present in the response latencies (see Kaiser et al., 2000; Krumbholz et al., 2007), and were thus not detectable in the current fMRI study.

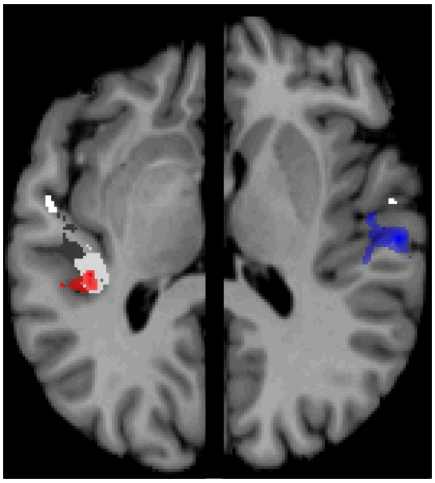

FIG. 3.4 Comparison of motion-right (red) and motion-left (blue) contrast images with the respective right-left flipped contrasts. As both the motion-right and motion-left conditions resulted in a predominantly contralateral response, only the contralateral hemisphere is shown for each condition. The oblique axial slice (parallel to the Sylvian fissure as in Fig. 3.3) was chosen to include the most significant voxel in each comparison.

3.D CONCLUSIONS

Overall, the current results show that the continuous stimulation paradigm (CSP) can be used with rapid event-related fMRI to measure transient feature-specific responses in auditory cortex with high spatial resolution. In particular, the results indicate that transient feature-specific responses exhibit a similar topography and hemispheric distribution as the sustained responses measured in epoch-related designs. This would strongly suggest that the EEG responses measured with the CSP (see **Chapter 2**) were feature-specific and not merely unspecific change responses.

While event-related fMRI designs are generally less efficient than epoch-related designs (Friston et al., 1999; Liu et al., 2001), the CSP would appear to have several distinct advantages. First, by avoiding the need for subtraction, the CSP would be expected to be considerably more powerful than other event-related designs that do involve subtraction. Second, adaptation paradigms in general, and the CSP in particular, have been shown to be more sensitive than epoch-related designs when comparing responses to features or feature values that would not be spatially resolvable with standard fMRI methods, such as orientation tuning in the visual cortex (Tootell et al., 1998; for review, see Grill-Spector et al., 2006). Third, unlike epoch-related designs, event-related paradigms such as the CSP open the possibility to consider the effects of perceptual factors, such as response time or response accuracy, on sensory-driven responses (Buckner et al., 1996). Finally, the CSP would seem an excellent candidate for combining fMRI and time-sensitive electrophysiological methods (EEG and MEG), and in particular for simultaneous EEG-fMRI recordings (for review, see Herrmann and Debener, 2008).

Chapter 4. Does binaural sluggishness affect pitch processing in binaurally unmasked low-frequency pure tones?[4]

4.A INTRODUCTION

Auditory-nerve responses mediate two types of temporal information, which can loosely be related to two different time scales. The rapidly-varying information, referred to as temporal fine-structure information (see **General introduction**), is associated with the individual cycles of the basilar membrane vibration at a given point and is mediated by phase locking. The more slowly-varying information, referred to as temporal envelope information, is related to the time-varying amplitude of this vibration. Temporal envelope information underlies the pitch of spectrally unresolved harmonic tones and amplitude-modulated noise. The excitation pattern of these stimuli contains no resolved harmonic peaks, and the temporal fine-structure of their auditory-nerve responses is largely determined by each fibre's characteristic frequency and thus conveys little or no information about the stimulus (Carney and Yin, 1988). In contrast, the auditory-nerve responses to low-frequency pure tones provide temporal fine-structure but no envelope information. Importantly, pure tones also convey spectral cues to pitch, as their excitation pattern contains a single peak, the location of which is related to the pure-tone frequency. Since both types of information are present, it is difficult to determine whether pitch perception in low-frequency pure tones is based on one or the other type of information.

This question has been a topic of debate for over a century (Seebeck 1843; Helmholtz 1863). At present, the dominant view appears to be that the pitch of low-frequency pure tones is based on temporal information (for review, see Moore, 2003, 2008; Plack and Oxenham, 2005). The most-cited argument in favour of this view is the

[4] Based on Magezi et al. (2009b)

observation that the accuracy of pure-tone frequency discrimination declines sharply above about 4 kHz (Moore, 1973) and the fact that this decline coincides with the assumed upper limit of phase locking in humans, which is derived from animal data (Johnson, 1980). However, contrary to this view, it has been argued that the decline in frequency discrimination accuracy above 4 kHz may also be due to a lack of selective pressure for humans to sustain accurate frequency discrimination at higher frequencies, because behaviourally relevant sounds such as speech and music contain most of their energy at lower frequencies (Heffner et al., 2001a; see also Sivian et al., 1959; Byrne et al., 1994).

In order to avoid this potential confound in comparing low and high frequencies, it would be desirable to be able to eliminate pitch-related temporal information from the internal neural representation of low-frequency sounds and determine the effect that this has on pitch perception at low frequencies. Recent results by Krumbholz et al. (2009a) suggest that binaural unmasking might constitute such a condition. They measured amplitude modulation (AM) detection and pitch-interval recognition thresholds for noise and harmonic-tone signals, respectively, when the signals were presented antiphasically (interaural phase difference of 180º or π) in a diotic noise masker. At low frequencies, antiphasic presentation of a signal in a diotic masker (often referred to as N_0S_π) results in a substantial reduction in detection threshold compared to homophasic (diotic) presentation (N_0S_0). This phenomenon is known as binaural unmasking (for review, see Durlach and Colburn, 1978). The antiphasic signals were presented at levels below their homophasic detection thresholds, and were thus perceived only through binaural channels. This was important, because the internal representation of a signal presented in

such conditions is known to be temporally smeared by binaural sluggishness (Hall and Grose, 1992; Culling and Colburn, 2000; van der Par et al., 2005). The results from the AM detection experiment in Krumbholz et al.'s (2009) study showed that, due to binaural sluggishness, the internal representation of binaurally unmasked sounds conveys little or no temporal envelope information at rates within the pitch range, i.e., above about 30-40 Hz (Krumbholz et al., 2000; Pressnitzer et al., 2001). Moreover, the results from the pitch-interval recognition experiment indicated that spectrally unresolved harmonic tones, the pitch of which relies exclusively on temporal envelope cues, fail to elicit pitch when presented in conditions of binaural unmasking. Binaural sluggishness has been modelled as a moving-average filter, which integrates the instantaneous output of the binaural processor according to a temporal weighting function, referred to as the "binaural temporal window", with a duration of the order of several tens to a few hundreds of milliseconds (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999). Binaural sluggishness would therefore be expected to smear the faster-varying temporal fine-structure information to an even greater extent than the more slowly varying temporal envelope information. If the large performance difference in frequency discrimination between low- and high-frequency pure tones reported in the previous literature (Moore, 1973) is due to the availability of temporal fine-structure information at low but not at high frequencies, then frequency discrimination performance for binaurally unmasked low-frequency pure tones would be expected to resemble that observed for high-frequency pure tones in normal listening conditions. The current study comprised three experiments aimed at testing this hypothesis by applying similar stimulus and task paradigms as have

been used for comparing frequency discrimination accuracy at low and high frequencies (Moore and Glasberg, 1989; Moore and Sek, 1994,1996) to low-frequency pure tones presented in conditions of binaural unmasking and in comparable diotic (homophasic) masking conditions.

4.B EXPERIMENT 1

4.B.i INTRODUCTION

Models that combine methods from signal detection theory with stochastic simulations of auditory-nerve activity predict that pure-tone frequency discrimination thresholds should be two orders of magnitude larger when only spectral (or "rate-place") information is taken into account compared to when temporal fine-structure information is also considered (Siebert, 1970; Heinz et al., 2001a). Furthermore, these models also predict a difference in the effect of stimulus duration on frequency discrimination performance. In particular, frequency discrimination thresholds would be expected to vary with the square root of the stimulus duration, $\sqrt{T}$, if discrimination were based on spectral information only, but with $\sqrt{T^3}$ if discrimination were based mainly on temporal information. Moore's (1973) perceptual data are at least in qualitative agreement with these predictions, in that they show that frequency discrimination thresholds at 8 kHz are about an order of magnitude larger than at 2 kHz (compare open and filled symbols in Fig. 4.1a), and the threshold-duration functions become progressively shallower above about 4 kHz (Fig. 4.1b).

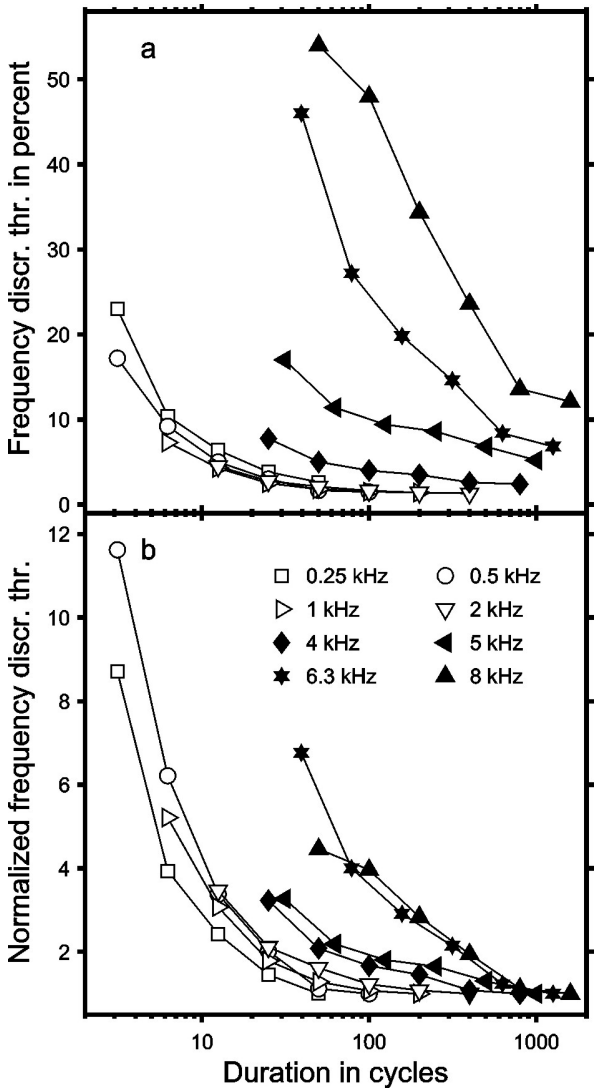

FIG. 4.1 Frequency discrimination thresholds for pure tones with different frequencies (denoted by different symbols; see legend), plotted as a function of stimulus duration. Thresholds are expressed in percent of the nominal frequency in panel a, and normalized to the asymptotic threshold at the longest duration tested for each frequency in panel b. Data are for one participant, replotted from Moore (1973). Frequencies that are thought to be coded temporally are denoted by open symbols, and frequencies that are thought to be coded spectrally by filled symbols.

In the first experiment of the current study, pure-tone frequency discrimination threshold was measured as a function of stimulus duration when the pure tones were presented in conditions of binaural unmasking or in comparable diotic masking conditions. The rationale was that, if the pitch of low-frequency pure tones relied on temporal fine-structure information, and if binaural sluggishness eliminated this information as it has recently been shown to eliminate pitch-related temporal envelope information (Krumbholz et al., 2009a), frequency discrimination threshold would be expected to be by at least an order of magnitude larger for the binaurally unmasked tones than for the diotically masked tones, and the function relating frequency discrimination threshold to stimulus duration would be expected to be considerably shallower.

4.B.ii METHODS

Stimuli

As in **Chapter 1** (Sec. 1.B), stimuli were generated digitally with a sampling rate of 25 kHz and a 24-bit amplitude resolution using TDT System 3 (Tucker-Davies Technology, Alachua, FL) and MATLAB® (The Mathworks, Natick, MA). They were digital-to-analogue converted (TDT RP2.1), passed through a headphone amplifier (TDT HB7) and presented via headphones (K240 DF, AKG, Vienna, Austria) to the participant, who was seated in a double-walled sound-attenuating room.

A noise masker was presented continuously and identically to both ears. The noise was filtered so as to produce an approximately constant level of excitation per equivalent rectangular bandwidth (ERB; Glasberg and Moore, 1990) and was presented at a fixed level of 55 dB SPL per ERB. The tones were presented either antiphasically (N_0S_π) or homophasically (N_0S_0) in this diotic noise masker.

Two pure-tone frequencies of 350 and 750 Hz were used. The starting phase of the tones was randomized for each presentation. Frequency discrimination thresholds were measured at five durations, corresponding to 2.5, 5, 10, 40 and 160 cycles of the tones' repetition period (about 2.86 and 1.43 ms for 350 and 700 Hz, respectively). The duration included squared-cosine gates of 1.25 cycles. This meant that the tones with the shortest duration (2.5 cycles) contained no steady-state portion.

In order to set the sensation level (SL) of the tones such that they would be well audible but, at the same time, the level of the antiphasic tones would not exceed the homophasic detection threshold, the detection threshold of the tones was measured in both the homophasic and antiphasic masking conditions for the median duration of 10 cycles. Previous research indicates that the masking level difference between the antiphasic and homophasic conditions (referred to as the binaural masking level difference, or BMLD) may depend slightly on signal duration (Blodgett et al., 1958; Green, 1966; Robinson and Trahiotis, 1972). In order to test whether this was a factor in the current experiment, the detection threshold was also measured for the longest duration of 160 cycles. In the frequency discrimination measurements, tones of different durations were presented with a constant overall energy, corresponding to 10 dB above the average of the detection thresholds at 10 and 160 cycles for the respective frequency and masking condition. In Fig. 4.2, the energy of the tones at detection threshold is expressed in terms of the equivalent level in dB SPL of an ungated 1-kHz tone with a duration of 10 ms (10 cycles).

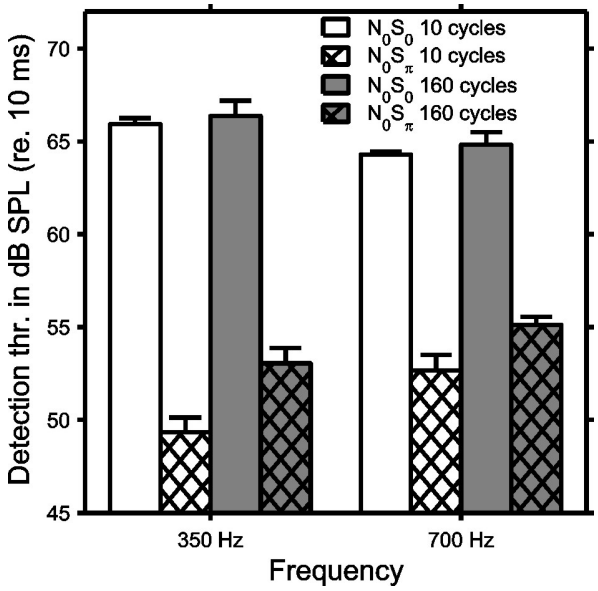

FIG. 4.2 Detection thresholds for homophasic (N_0S_0, open bars) and antiphasic (N_0S_π, hatched bars) pure tones in a continuous, diotic noise masker for different tone frequencies [350 (left set of bars) and 700 Hz (right set of bars); see abscissa] and durations [10 (bars with white background) and 160 cycles of the repetition period of the tones (bars with grey background) see legend]. Thresholds were averaged over four participants and expressed in terms of the equivalent level (in dB SPL) of an ungated 1-kHz tone with a 10-ms duration. Error bars show standard errors.

Procedure

Both the detection and frequency discrimination thresholds were measured with an adaptive two-interval, two-alternative forced-choice procedure (2I2AFC), using a three-down one-up rule, which tracks the 79% correct point on the psychometric function (Levitt, 1971). Each trial consisted of two observation intervals of 250 ms, separated by a gap of 500 ms.

In the detection measurements, only one of the intervals contained a tone. The task was to identify that interval using a button box. Visual feedback was provided at the end of each trial. The level of the tone was changed in steps of 5 dB up to the first reversal in level, 3 dB up to the second reversal and 2 dB for the rest of the 12 reversals that made up each threshold run. The arithmetic mean of the levels at the last 10 reversals was calculated to obtain a single threshold estimate. At least three such threshold estimates were averaged to obtain the final threshold estimate for each condition.

In the frequency discrimination measurements, both intervals contained a tone, and the frequency of the tones differed by a percentage, ΔF, of their mean. The task was to identify the interval containing the higher-frequency tone. The frequency difference between the two tones was reduced and increased by a factor, v. Up to the first reversal, v was equal to 1.9; v was reduced to 1.5 up to the second reversal and set to 1.3 for the following 10 reversals. The geometric mean of the percentage frequency differences for the last 10 reversals was calculated to obtain a single threshold estimate. The final threshold estimate for each condition was the mean of at least three such threshold estimates. In order to minimize the potential for loudness cues, the mean frequency for each trial was randomized by up to 3% around the nominal frequency of 350 or 700 Hz.

For both the detection and frequency discrimination measurements, the order in which different conditions were tested was randomized for each of the three threshold runs.

Data analysis

The average detection thresholds of each participant were submitted to a three-way repeated-measures analysis of variance (ANOVA), with factors frequency (350 and 700 Hz), masking condition (N_0S_0 and N_0S_π) and tone duration (10 and 160 cycles).

The ANOVA assumes equality ("homogeneity") of the variances of the different groups of data (see, e.g., Howell, 2002). Many previous studies have converted frequency discrimination or frequency modulation (FM) detection thresholds into logarithmic units, even when the thresholds were expressed in percent as in the current study (e.g., Moore and Sek, 1996; Micheyl et al., 1998), because threshold variances were found to be heterogeneous on a linear scale. In the current study, frequency discrimination thresholds were evaluated in both linear and logarithmic units, and the degree of heterogeneity of variance was analyzed by computing the correlation between the inter-participant means and variances of the thresholds in both linear and logarithm units. While there was a strong positive correlation for the linear thresholds ($r = 0.637$, $p = 0.003$), there was an equally strong, but negative correlation for the logarithmic thresholds ($r = –0.686$, $p = 0.001$). Both the linear and logarithmic thresholds were submitted to a three-way repeated-measures ANOVA with factors frequency (350 and 700 Hz), masking condition (N_0S_0 and N_0S_π), and tone duration (2.5, 5, 10, 40 and 160 cycles).

Participants

Four participants (3 male and 1 female, aged between 20 and 21 years) were tested. All participants in this study (including those who took part in the other two experiments presented in this study) had absolute thresholds of less than 25 dB HL at audiometric frequencies, and had no history of hearing or neurological disorders. They either had previous experience in psychoacoustic experiments or were lay musicians. Participants who were not authors of the corresponding manuscript (Magezi et al., 2009b) were paid for their services at an hourly rate. The experimental procedures were approved by the Ethics Committee of the Nottingham University School of Psychology.

4.B.iii RESULTS

The detection thresholds for both frequencies (350 and 700 Hz) and tone durations tested (10 and 160 cycles) showed a substantial masking release in the antiphasic compared to the homophasic masking condition (compare open and hatched bars in Fig. 4.2). The BMLD amounted to 12.8 (±1.5) dB on average and was significant [main effect of masking condition: $F(1,3) = 254.762$, $p = 0.001$]. The BMLD was significantly larger (4.3 dB) for 350 than for 700 Hz [interaction between masking condition and frequency: $F(1,3) = 200.359$, $p = 0.001$], which is consistent with previous data (Durlach and Colburn, 1978). The shorter tones (10 cycles) were detected at slightly lower energy levels than the longer tones (160 cycles) [main effect of duration: $F(1,3) = 11.081$, $p = 0.045$]. Moreover, the effect of duration was more prominent for the antiphasic than the homophasic tones [interaction between masking condition and duration: $F(1,3) = 21.747$, $p = 0.019$], which meant that the BMLD was on average 2.6 dB larger for the shorter than the longer tones (compare bars with white and grey backgrounds in Fig. 4.2). This finding is consistent with previous data (Blodgett et al.,

1958; Green, 1966, Robinson and Trahiotis, 1972). Neither the main effect of frequency, nor the interactions with frequeny (interaction between frequency and duration, and the three-way interaction) were significant.

As expected from previous research (Turnbull, 1944; König, 1957; Liang and Chistovich, 1961; Sekey, 1963; Henning, 1970; Moore, 1973; Hall and Wood, 1984; Freyman and Nelson, 1986), frequency discrimination performance showed a substantial improvement with increasing tone duration (Fig. 4.3), yielding significant main effects of duration in both the linear [$F(4,12) = 49.121$, $p < 0.001$] and logarithmic threshold data [$F(4,12) = 27.729$, $p < 0.001$]. The improvement was limited to tone durations of up to 10 cycles, where the threshold functions started to reach an asymptote.

Based on the assumption that pitch perception in low-frequency pure tones is based on temporal fine-structure information and that binaural sluggishness eliminates this information in conditions of binaural unmasking, we expected the frequency discrimination threshold to be substantially larger, and the slope of the function relating threshold to duration to be considerably shallower, for the antiphasic than the homophasic condition. While there was a small tendency for threshold to be larger (Fig. 4.4a), and the threshold function to be shallower (Fig. 4.4b), in the antiphasic than the homophasic condition, these differences were much smaller than expected and only reached significance for the logarithmic [main effect of masking condition: $F(1,3) = 15.902$, $p = 0.028$; interaction between masking condition and duration: $F(4,12) = 4.373$, $p = 0.021$] but not the linear data [main effect: $F(1,3) = 1.589$, $p = 0.297$; interaction: $F(4,12) = 0.204$, $p = 0.931$]. Neither the the main effect of frequency, nor the interactions with frequency (frequency and masking condition, frequency and duration, the three-way

interaction) were significant in either logarithmic or linear units. Collapsing the data across frequency (Fig. 4.4a) showed that the asymptotic threshold for the antiphasic condition (average across durations greater than or equal to 10 cycles) was about 1.5 times larger than for the homophasic condition (corresponding to a linear threshold difference of only 0.8%). This difference is consistent with previous studies (Henning, 1990; Henning and Wartini, 1990; Hall et al., 1997) and is almost an order of magnitude smaller than the difference in frequency discrimination threshold found between high and low frequencies in normal listening conditions (compare Figs 4.1 and 4.4), and two orders of magnitude smaller than the difference predicted by models based on temporal or spectral information in the auditory nerve (Siebert, 1970; Heinz et al., 2001a). Rather than reflecting a difference in processing mechanism, the small difference in asymptotic frequency discrimination threshold could be related to the difference in loudness between homophasic and antiphasic tones presented at the same sensation level (Townsend and Goldstein, 1970; Soderquist and Shilling, 1990), and the difference in the effect of duration on the threshold functions may be due to the fact that the BMLD was slightly larger at shorter than at longer durations. This idea is supported by results of Henning and Wartini (1990), who measured frequency discrimination thresholds for homophasic and antiphasic tones at three durations ranging from 14-47 cycles and failed to find any difference in the effect of duration between the masking conditions.

The results from the current experiment suggest that frequency discrimination performance for pure tones presented in conditions of binaural unmasking is remarkably similar to that for pure tones presented in comparable diotic masking conditions, suggesting that pitch is processed by the same mechanism in both cases. In order to

investigate this conclusion further, Experiments 2 (Sec. 4.C) and 3 (Sec. 4.D) were aimed at measuring the effect of disrupting spectral pitch cues in pure tones presented in homophasic and antiphasic masking conditions, using similar paradigms as those developed by Moore and colleagues to investigate pitch mechanisms in normal listening conditions (Moore and Glasberg, 1989; Moore and Sek, 1996).

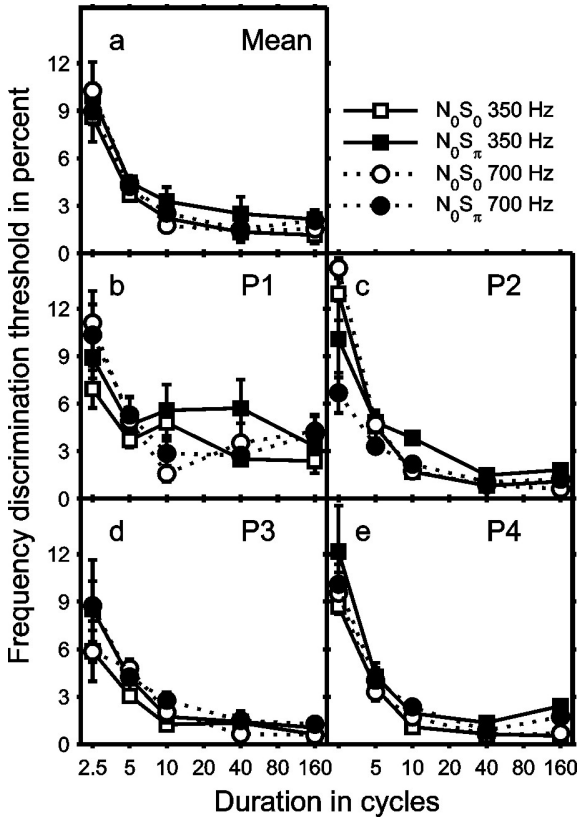

FIG. 4.3 Average (a) and individual (b – e) frequency discrimination thresholds plotted as a function of tone duration in cycles of the repetition period. Thresholds are expressed in percent of the nominal frequency. Open and filled symbols show thresholds measured in the homophasic (N_0S_0) and antiphasic (N_0S_π) masking conditions, respectively (see legend). Thresholds for 350 and 700 Hz are denoted by squares and circles, respectively (legend). Error bars show standard errors.

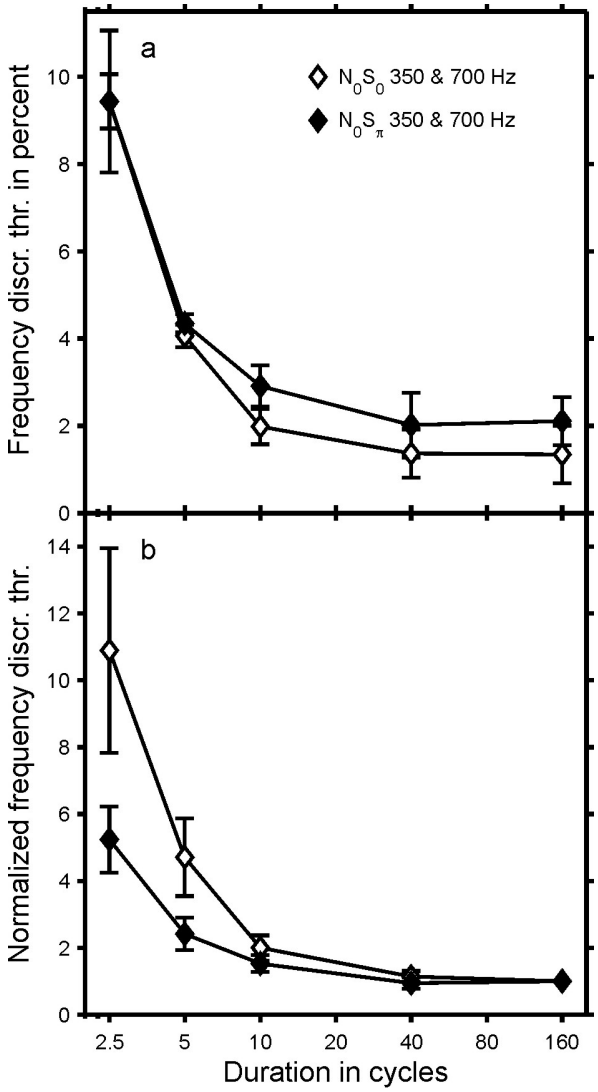

FIG. 4.4 Average frequency discrimination thresholds collapsed across frequencies (350 and 700 Hz). Thresholds are expressed in percent of the nominal frequency in panel a, and normalized to the asymptotic threshold at the longest duration of 160 cycles in panel b. As in Fig. 4.3, the abscissae show tone duration in cycles of the repetition period, and thresholds for the homophasic (N_0S_0) and antiphasic (N_0S_π) masking conditions are denoted by open and filled symbols, respectively.

4.C EXPERIMENT 2

4.C.i INTRODUCTION

In a series of studies, Moore and Glasberg (1989) and Moore and Sek (1996, 1998) devised different methods of disrupting spectral pitch cues by introducing random level changes and investigated the impact that this had on frequency discrimination performance. In one experiment, Moore and Sek (1996) measured the detectability of sinusoidal frequency modulation (FM) in pure-tone carriers, the amplitude of which was also modulated over time to disrupt spectral FM cues (Fig. 4.5). Moore and Sek found that, at slow modulation rates (below 5 Hz), the AM had a large detrimental effect on FM detection at high frequencies, but only had a relatively small effect at low frequencies, and they suggested that this was due to the availability of temporal pitch cues at low frequencies and slow FM rates.

In the current study, Moore and Sek's (1996) FM detection paradigm was applied to pure tones presented in conditions of binaural unmasking. FM detection thresholds were measured with and without concurrent AM (labelled "AM" and "no AM", respectively) for pure tones presented either homophasically or antiphasically in a diotic noise masker. For comparison, FM thresholds with and without concurrent AM were also measured for tones presented in quiet (labelled "no noise"). If binaural sluggishness smeared temporal fine-structure information in binaurally unmasked sounds, FM perception in the antiphasic masking condition would have to be based purely on spectral cues. In view of Moore and Sek's results, it was thus expected that the concurrent AM would have a detrimental effect on FM detection performance in the antiphasic but not the homophasic masking condition.

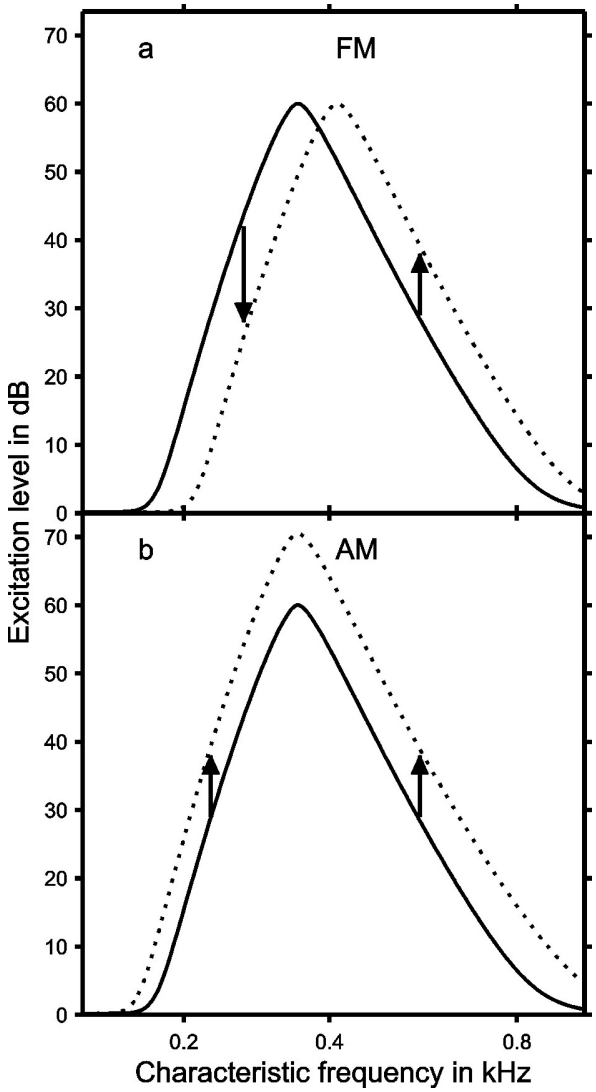

FIG. 4.5 Illustration of the changes in a pure tone's excitation pattern when the frequency (a) or amplitude (b) of the tone is changed. The arrows show that both the frequency and the amplitude change produce changes in excitation level along the flanks of the pattern. Zwicker (1956, 1970) assumed that both FM and AM detection are based on these excitation-level changes. According to this model, FM detection would thus be expected to be disrupted by concurrent AM.

4.C.ii METHODS

Stimuli

Frequency-modulated tones were generated according to the general formula $s(t) = A \cdot \cos(2 \cdot \pi \cdot F_{\text{car}} \cdot t + \phi_{\text{car}} + \frac{\Delta F_{\text{c}}}{2 \cdot F_{\text{mod}}} \cdot \sin(2 \cdot \pi \cdot F_{\text{mod}} \cdot t + \phi_{\text{FM}}))$, where F_{car} is the carrier frequency of the tone, which was 350 Hz throughout this experiment, ΔF_{c} is the frequency excursion of the FM (difference between maximal and minimal frequency), F_{mod} is the modulation rate, and ϕ_{car} and ϕ_{FM} are the starting phases of the carrier and the FM, respectively. FM detection thresholds were measured for four different modulation rates, F_{mod}, of 2, 5, 10 and 20 Hz. In the AM conditions, the amplitude, A, was a function of time: $A(t) = 1 + m \cdot \cos(2 \cdot \pi \cdot F_{\text{mod}} \cdot t + \phi_{\text{AM}})$, where m is the modulation index and ϕ_{AM} is the starting phase of the AM. The AM modulation depth was fixed at m = 0.332 (corresponding to a peak-to-trough ratio of 6 dB) as in the studies by Moore and Glasberg (1989) and Moore and Sek (1996). A 6-dB modulation depth is thought to be large enough to measurably disrupt spectral FM cues whilst minimizing level-related pitch shifts. All three starting phases (ϕ_{car}, ϕ_{FM} and ϕ_{AM}) were randomized for each presentation. The tones had an overall duration of 1000 ms, including 15-ms squared-cosine ramps.

Prior to the FM detection measurements, the masked detection threshold of an unmodulated tone (containing neither FM nor AM) was measured in homophasic and antiphasic conditions, and the levels of the tones in the FM detection measurements were set to 10 dB above the respective detection thresholds. The tones presented in quiet were set to the same level in dB SPL as the tones in the homophasic condition. As in

Experiment 1 (Sec. 4.B), the masking noise was an equally-exciting noise, which was presented continuously at an ERB level of about 55 dB SPL. The set up and general aspects of the stimulus generation were the same as in Experiment 1 (see Sec. 4.B.ii).

Procedure

The detection and FM thresholds were measured with a similar three-down one-up 2I2AFC adaptive procedure as used in Experiment 1 (Sec. 4.B.ii). The observation intervals were 1000 ms in duration, separated by a 500-ms gap. In the FM detection measurements, participants were asked to indicate which of two tones was frequency modulated, and to ignore the AM when it was present. The adaptive parameter was the excursion, ΔF_c, of the FM. ΔF_c was varied in the same multiplicative way as the frequency difference, ΔF, in the frequency discrimination measurements of Experiment 1. Other aspects of the procedure were the same as in Experiment 1 (Sec. 4.B.ii).

Data analysis

The detection thresholds for the homophasic and antiphasic conditions were compared by means of a paired *t*-test. As for the frequency discrimination thresholds measured in Experiment 1, the FM detection thresholds were evaluated in both linear and logarithmic units (see "Data analysis" in Sec. 4.B.ii) and submitted to a three-way repeated-measures ANOVA with factors masking condition (N_0S_0, N_0S_π and no noise), AM condition (AM and no AM) and modulation rate (2, 5, 10 and 20 Hz). In addition, the FM detection thresholds for the no-noise condition were submitted to a two-way repeated-measures ANOVA with factors AM condition (AM and no AM) and modulation rate (2, 5, 10 and 20 Hz). As in Experiment 1 (Sec. 4.B.ii), a correlation analysis between the means and variances of the FM detection thresholds showed that,

while there was a strong positive correlation for the linear thresholds (r = 0.481, p = 0.017), there was an equally strong, but negative correlation for the logarithmic thresholds (r = -0.521, p = 0.009).

Participants

Four participants were tested (1 male and 3 female, aged between 19 and 23 years), of whom one had also participated in Experiment 1 (Sec. 4.B.ii).

4.C.iii RESULTS

The masking level difference for the unmodulated 350-Hz tone between the antiphasic and homophasic conditions amounted to 13.8 (±0.6) dB and was highly consistent across participants [$t(3) = 22.064$, $p < 0.001$].

Figure 4.6 shows that FM detection performance strongly depended on masking condition [main effect of masking condition (linear/logarithmic): $F(2,6)$ = 100.015/83.859, $p < 0.001/0.001$], with the lowest and highest FM detection thresholds in quiet (squares) and in the antiphasic condition (triangles), respectively, and intermediate thresholds in the homophasic condition (circles). There was also a general increase in FM detection threshold with increasing modulation rate [main effect of modulation rate (linear/logarithmic): $F(3,9)$ = 51.161/73.416, $p < 0.001/0.001$]. The main effect of AM condition was statistically significant [linear/logarithmic: $F(1,3)$ = 50.223/318.540, p = 0.006/0.001], even though only the no-noise condition contributed to this effect [interaction between masking condition and AM condition (linear/logarithmic): $F(2,6)$ = 5.666/59.733 , p = 0.041/0.001]. In the no-noise condition, FM detection was practically unaffected by AM at the lowest FM rate of 2 Hz. This finding was attributed by Moore and Sek (1996) to the availability of temporal pitch information. At higher FM rates, AM

had an increasingly detrimental effect on FM detection [filled squares in Fig. 4.6; three-way interaction (linear/logarithmic): $F(6,18) = 3.624/8.341$, $p = 0.016<0.001$]. An ANOVA of the no-noise condition revealed significant main effects and interaction [main effect of AM condition (linear/logarithmic): $F(1,3) = 55.327/226.212$, $p = 0.005/0.001$; main effect of modulation rate: $F(3,9) = 6.237/10.521$, $p = 0.014/0.03$; interaction : $F(3,9) = 5.004/6.310$, $p = 0.026/0.014$]. Moore and Sek explained this finding by assuming that the temporal pitch mechanism is unable to follow faster changes in frequency, so that FM perception becomes progressively more reliant on spectral cues at faster modulation rates.

Based on Moore and Sek's (1996) results and the assumption that temporal fine-structure information is eliminated by binaural sluggishness in conditions of binaural unmasking, we expected FM detection performance in the antiphasic condition to show a significant effect of AM even at the slowest modulation rate of 2 Hz. In contrast, the pattern of FM detection thresholds for the homophasic condition was expected to resemble that for the no-noise data, with a significant effect of AM condition at the faster modulation rates, but little or no effect at 2 Hz. Unfortunately for the sake of the current argument, the AM (open symbols in Fig. 4.6) and no-AM conditions (filled symbols) yielded practically identical FM detection thresholds at all modulation rates in the homophasic and antiphasic masking conditions [interaction between modulation rate and AM condition $F(3,9) = 0.572/1.665$, $p = 0.647/0.243$], suggesting that the amount of AM used was too small to affect FM detection performance in these conditions, so that its effect was overridden by other factors. In the homophasic condition (circles in Fig. 4.6), FM detection performance was probably limited by the intrinsic fluctuations in the

masker. Due to these intrinsic fluctuations, the tone level above the masker would fluctuate in a similar way as the level of the amplitude-modulated tone in the no-noise condition, explaining why there was practically no added effect of AM in the homophasic condition. The FM detection thresholds for the antiphasic tones (triangles in Fig. 4.6) were not only considerably larger than the thresholds for homophasic tones, but also exhibited a steeper initial increase with increasing modulation rate [interaction between masking condition and modulation rate (linear/logarithmic): $F(6,18) = 10.948/4.680$,$p < 0.001/ = 0.05$]. The fact that the functions levelled off towards higher rates (10 and 20 Hz) was probably due to spectral sidebands becoming audible in these conditions. The larger FM detection thresholds in the antiphasic condition, and the steeper initial increase in threshold with FM rate, are both likely to be a consequence of binaural sluggishness. Irrespective of what kind of cues pitch is based on (spectral or temporal), binaural sluggishness would be expected to smooth the temporal changes in these cues introduced by the FM, making the FM harder to detect (Culling and Colburn, 2000). Under the assumption that, at the lowest rate of 2 Hz, FM detection in the homophasic condition was limited by a leaky integrator with a 2.5-ms time constant (Green, 1973; Viemeister, 1979), the binaural integrator would have to have a time constant of about 150 ms to explain the FM threshold in the antiphasic condition at 2 Hz, which was more than twice as large as the homophasic threshold. This estimate is consistent with previous estimates of the integration time constant underlying binaural sluggishness based on detection data (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999). The fact that, as for the homophasic condition, AM had little effect on FM detection performance in the antiphasic condition (compare

open and filled triangles) was probably due to the antiphasic FM thresholds being too high even in the no-AM condition for the 6-dB AM to have any appreciable effect. Moreover, in the antiphasic condition, the effect of AM would be expected to have been reduced by binaural sluggishness (Hall and Grose, 1992; Krumbholz et al., 2009a).

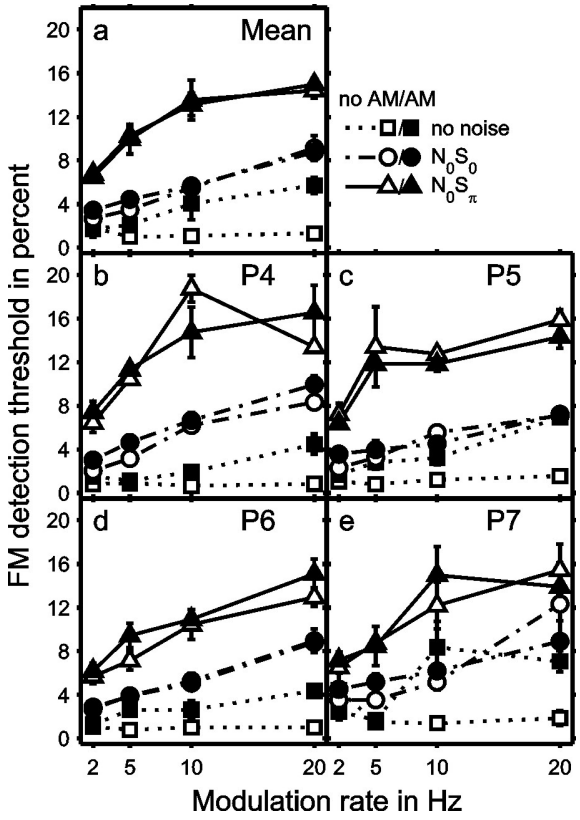

FIG. 4.6 Average (a) and individual (b – e) FM detection thresholds plotted as a function of the modulation rate in Hz. Thresholds are expressed as the frequency excursion of the FM (maximum minus minimum frequency) in percent of the mean frequency. Different masking conditions are depicted by different symbols and line types (no-noise: squares; N_0S_0: circles; N_0S_π: triangles) and AM condition is represented by marker colour (no AM: open; AM: filled). Error bars show standard errors.

The data from the current experiment were inconclusive with respect to the effect of AM on FM detection in conditions of binaural unmasking, because the amount of AM used was too small to have an effect in any but the no-noise condition. In order to address this problem, we conducted another experiment using a paradigm developed by Moore and Sek (1992), which combines AM and FM of equal perceptual salience.

4.D EXPERIMENT 3

4.D.i INTRODUCTION

In Experiment 2 (Sec. 4.C), the amount of AM used had turned out insufficient to appreciably disrupt spectral FM cues. In Experiment 3 (current section), another of the paradigms developed by Moore and Sek was used, which combines AM and FM with equal or similar perceptual salience. Moore and Sek (1992; see also Moore and Sek, 1994, 1996) measured the detectability of combined AM and FM, referred to as mixed modulation (MM), as a function of the relative starting phases of the two modulators. When the difference between the starting phases of the AM and FM is zero, the amplitude maxima coincide with the frequency maxima; when the difference is 180º or π, the amplitude maxima coincide with the frequency minima. In MM detection, both the FM and the AM are only presented in the signal interval so they both serve as potential cues for the modulation detection task. Moore and Sek argued that, in conditions where both AM and FM are based on spectral information, MM detection performance should depend on the phase difference between the two modulators. If both sides of the excitation pattern were equally important for MM detection, phase differences of zero and π would be expected to yield the best possible performance, because, for these phase

differences, the AM- and FM-related excitation-level changes summate on one side of the pattern (see Fig. 4.5 and Moore and Sek, 1992, 1994). Intermediate phase differences of $\pi/2$ or $3\pi/2$, on the other hand, would be expected to yield the worst performance. Moore and Sek (1996) found that, at high frequencies, where both AM and FM are thought to be coded spectrally, relative modulator phase did indeed have a significant effect on MM detection performance. However, the pattern of results differed somewhat from the above-mentioned expectations. In particular, Moore and Sek found that MM detection performance was best for zero phase difference, intermediate for $\pi/2$ and $3\pi/2$, and worst for π, suggesting that MM detection in these conditions was mainly based on the high-frequency side of the excitation pattern, where AM- and FM-related excitation-level changes summate when the modulator phase difference is zero, and cancel when the phase difference is π (see Fig. 4.5). Importantly, at low frequencies and slow modulation rates, where FM is thought to be coded temporally, Moore and Sek (1996) found MM detection performance to be independent of the modulator phase difference.

The current experiment measured MM detection at a low frequency (350 Hz) and a slow modulation rate (2 Hz). As in Experiments 1 (Sec. 4.B) and 2 (Sec. 4.C), the tones were presented in homophasic and antiphasic masking conditions. Based on Moore and Sek's (1996) findings, MM detection performance was expected to be independent of the modulator phase difference in the homophasic condition. In the antiphasic condition, on the other hand, MM detection performance would be expected to show a similar pattern of phase effects as has previously been observed at high frequencies (Moore and Sek, 1996) if binaural sluggishness degraded temporal FM cues in this condition. In addition to the MM detection task, the current experiment also included an FM detection task

similar to that used in Experiment 2 (Sec. 4.C), but in this case using an AM distracter that was matched to the FM in terms of perceptual salience.

4.D.ii METHODS

Stimuli

Amplitude and frequency-modulated tones were generated according to the equations presented in the methods section of Experiment 2 (Sec. 4.C.ii). The carrier frequency (F_{car}) was 350 Hz and the modulation rate (F_{mod}) was fixed at 2 Hz. For the MM detection measurements, the phase difference between the amplitude and frequency modulators ($\Delta\phi = \phi_{AM} - \phi_{FM}$) was set to 0, $\pi/2$ or π. For the FM detection measurements, the starting phases of both modulators were randomized as in Experiment 2. As in Experiment 2, the levels of the tones in the homophasic and antiphasic masking conditions were set to 10 dB above the masked threshold of an unmodulated tone in the respective masking condition. The set up and other aspects of the stimulus generation were the same as in Experiment 2 (Sec. 4.C.ii).

Procedure

Masked detection thresholds were measured with a similar procedure as in Experiment 2 (Sec. 4.C.ii), except that each threshold run comprised of 10 rather than 12 reversals in level, and each threshold estimate was taken to be the arithmetic mean of the levels at the last 8 reversals. The MM detection task required the amounts of AM and FM to be matched in terms of their detectability when presented on their own. For that, thresholds for the detection of AM and FM alone were first obtained using an adaptive procedure. The FM detection thresholds were measured with the same procedure as used in Experiment 2 (Sec. 4.C.ii), except that each run comprised of 10 rather than 12

reversals. In the case of the AM detection thresholds, the adaptive parameter was the modulation depth, *m*, which was changed in logarithmic steps of 5 dB up to the first reversal, 3 dB up to the second reversal and 2 dB for the rest of the 10 reversals that made up each run. Each threshold estimate was taken to be the arithmetic mean of *m* in dB [$20 \cdot \log_{10}(m)$] at the last 8 reversals. At least three such threshold estimates were averaged to obtain the final threshold estimate for each condition. Then, psychometric functions were measured for the detection of AM and FM alone using the method of constant stimuli and the same 2I2AFC procedure as used for the adaptive threshold measurements. The adaptive threshold for each condition was used to select four modulation depths that would yield percent-correct performance levels ranging from chance (50%) to perfect (100%) performance. In most cases, these four modulation depths spanned a range of 10 dB, in the case of the AM, or 10%, for the FM, around the respective adaptive threshold. Data were collected in blocks, which comprised 20 presentations of each modulation depth. The masking condition (homophasic or antiphasic) and type of modulation (AM or FM) were kept constant within each block, and the four modulation depths were presented in a random order. Prior to each block, five practice trials were presented using the largest of the four modulation depths, where the modulation was most clearly audible. Four blocks were run for each condition yielding a total of 80 trials for each condition and modulation depth.

The psychometric functions for the detection of AM and FM alone were then used to derive the modulations depths for the MM and FM detection measurements. For that, a linear regression line was fitted to each function and used to select four modulation depths corresponding to percent-correct values that were equally spaced between 55%

and 95%. The resulting modulation depths for AM and FM alone were then combined to measure psychometric functions for MM detection and FM detection with and without concurrent AM. The procedure used to measure the psychometric functions for the MM and FM detection with and without AM was the same as that used to measure the psychometric functions for the detection of AM and FM alone.

For the FM detection with and without AM, only the largest three of the four modulation depths were used to ensure that the detectability of the FM was always well above chance level when no AM was added.

Data analysis

The detection thresholds for the homophasic and antiphasic conditions were compared by means of a paired *t*-test. For the AM- and FM-alone psychometric functions, the absolute values of the slopes of the linear regression lines were submitted to a two-way repeated-measures ANOVA with factors modulation type (AM and FM) and masking condition (N_0S_0 and N_0S_π). The percent-correct scores for the MM detection were submitted to a three-way repeated-measures ANOVA with factors masking condition (N_0S_0 and N_0S_π), modulator phase difference (0, $\pi/2$ and π) and modulation depth (four values). The percent-correct scores for FM detection with and without AM were also submitted to a three-way repeated-measures ANOVA with factors masking condition (N_0S_0 and N_0S_π), AM condition (AM and no AM) and modulation depth (three values).

Participants

Five participants were tested (1 male and 4 female, aged between 21 and 30 years). One participant only completed the MM detection measurements.

4.D.iii RESULTS

The BMLD for the unmodulated tone was similar to that found in Experiment 2 [13.1 (±0.8) dB; $t(4) = 16.594$, $p < 0.001$; see Sec. 4.C.iii].

Psychometric functions for AM, FM and MM detection were all monotonic, as expected (see Figs 4.7 and 4.8). The functions for the antiphasic masking condition were generally shifted towards higher modulation depths compared to the corresponding homophasic functions, indicating that modulation detection thresholds were generally larger in the antiphasic than in the homophasic condition (see, e.g., Fig. 4.7). Although this finding was not statistically significant [main effect of masking condition: $F(1,4) = 5.033$, $p = 0.088$], it is consistent with the findings from Experiment 2 (Sec. 4.C.iii) and from previous studies (Hall and Grose, 1992; Culling and Colburn 2000; Krumbholz et al., 2009a; see also Henning, 1990; Henning and Wartini, 1990) which have shown that the detection of changes in amplitude or frequency is degraded in antiphasic masking conditions, and that this is most likely due to binaural sluggishness. Interestingly, the slope of the psychometric function for the detection of FM alone was shallower for the antiphasic than for the homophasic condition (compare open and filled circles in Fig. 4.7). The same did not apply to the detection of AM alone [left panels in Fig. 4.7; main effect of modulation type: $F(1,4) = 50.936$, $p = 0.002$; interaction between modulation type and masking condition: $F(1,4) = 11.768$, $p = 0.027$]. We have currently no explanation for this difference.

Contrary to our expectations based on the assumption that temporal fine structure is not preserved in the internal representation of binaurally unmasked sounds, MM detection was unaffected by the phase difference between AM and FM for both the

homophasic and antiphasic masking conditions [Fig. 4.8; interaction between phase difference and masking condition: $F(2,8) = 1.082$, $p = 0.384$]. According to Moore and Sek's (1996) reasoning, this might be taken to suggest that FM was coded temporally in both conditions. However, as will be further discussed in the Discussion (Sec 4.E), the absence of an effect of relative modulator phase in the current as well as previous low-frequency data (Moore and Sek, 1996) may also be due to other reasons, and does thus not necessarily imply temporal coding of FM. As expected, there was a significant main effect of modulation depth [$F(3,12) = 265.662$, $p < 0.001$]. None of the other main effects or interactions was statistically significant.

The results for FM detection with concurrent AM appear to be in contrast with the MM detection results, in that FM detection performance was strongly degraded by the concurrent AM [compare circles and squares in Fig. 4.9; main effect of AM condition: $F(1,3) = 14.964$, $p = 0.031$]. For most participants, the detrimental effect of AM on FM detection performance increased with increasing modulation depth [main effect of modulation depth: $F(2,6) = 17.258$, $p = 0.003$; interaction between modulation depth and AM condition: $F(2,6) = 5.952$, $p = 0.038$]. The fact that the effect of AM was the same in both masking conditions [main effect of masking condition: $F(1,3) = 0.018$, $p = 0.901$; interaction between masking condition and AM condition: $F(1,3) = 0.471$, $p = 0.542$] would appear to suggest that FM was coded spectrally in both conditions. However, it is also possible that, as the signals in the current experiments had to be presented at a relatively low sensation level (10 dB SL; see Sec. 4.B.ii), they may have been rendered inaudible during the troughs of the AM. This would have made the FM more difficult to hear, irrespective of what cues its perception was based on. Neither the interaction

between masking condition and modulation depth [$F(2,6) = 0.365$, $p = 0.709$], nor the three-way interaction: $F(2,6) = 1.082$, $p = 0.397$] were significant. Taken together, the MM and FM detection results from the current experiment suggest that that FM is coded by the same mechanism in both homophasic and antiphasic masking conditions. Neither set of results seems to allow any definite conclusions as to whether that mechanism is based on spectral or on temporal information.

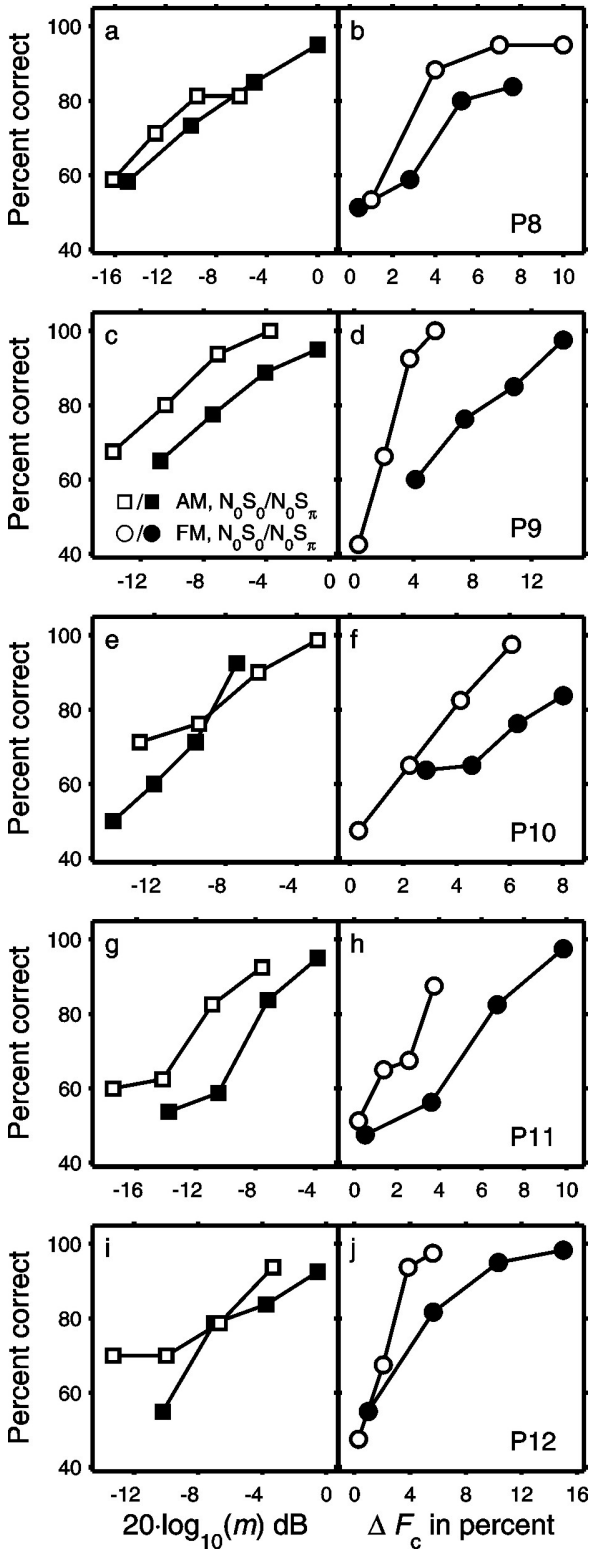
a
b
c
d
e
f
g
h
i
j
Percent correct
100
80
60
40
-16 -12 -8 -4 0
0 2 4 6 8 10
-12 -8 -4 0
0 4 8 12
-12 -8 -4
0 2 4 6 8
-16 -12 -8 -4
0 2 4 6 8 10
-12 -8 -4 0
0 4 8 12 16
□/■ AM, N_0S_0/N_0S_π
○/● FM, N_0S_0/N_0S_π
P8
P9
P10
P11
P12
20·log$_{10}$(*m*) dB
Δ F_c in percent

FIG. 4.7 Psychometric functions for the detection of AM (squares, left column) and FM (circles, right column). The AM modulation depth, *m*, is expressed in dB. The FM modulation depth is expressed as the frequency excursion, ΔF_c (maximum minus minimum frequency), in percent of the mean frequency. Open and filled symbols refer to homophasic (N_0S_0) and antiphasic (N_0S_π) conditions, respectively (see legend in panel c). Each row shows data from a different participant.

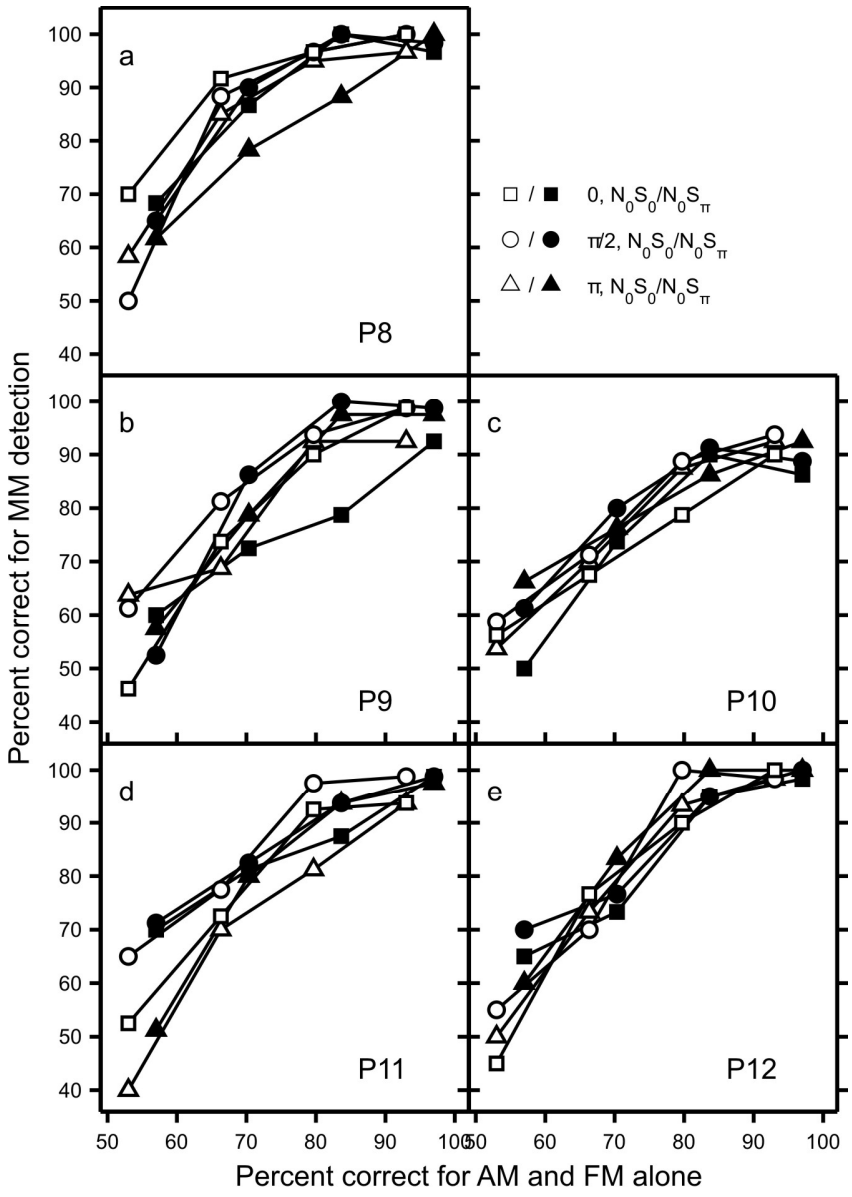

FIG. 4.8 Psychometric functions for the detection of mixed modulation (MM). The percent-correct scores for MM detection are plotted as a function of the nominal percent-correct scores for the detection of AM or FM alone. Different relative modulator phases are depicted by different symbols (0: squares; π/2: circles; π: upward pointing triangles), and masking condition is represented by marker colour (N_0S_0: open; N_0S_π: filled). Symbols are displaced along the abscissa for clarity.

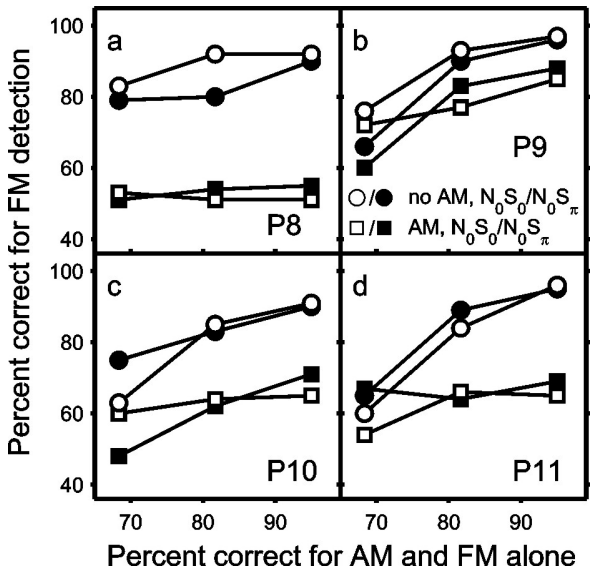

FIG 4.9 Psychometric functions for the detection of FM with and without concurrent AM. As in Fig. 4.8, percent-correct scores are plotted as a function of the nominal percent-correct scores for the detection of AM or FM alone. Open and filled symbols refer to homophasic (N_0S_0) and antiphasic (N_0S_π) masking conditions, respectively. The no-AM and AM conditions are denoted by circles and squares (see legend in panel b).

4.E DISCUSSION

This study was based on our previous finding that, in conditions of binaural unmasking, binaural sluggishness eliminates temporal envelope cues to pitch, with the striking perceptual consequence of precluding pitch perception in spectrally unresolved harmonic tones when presented in such conditions (Krumbholz et al., 2009a). The aim was to test whether binaural sluggishness also affects pitch-related temporal fine-structure information. Given that binaural sluggishness is thought of as a moving-average filter with a very long time constant (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999), the faster-varying fine-structure information would be expected to be smeared even more than the envelope information. Based on modelling predictions and previous perceptual findings, a smearing of temporal-fine structure information in binaural unmasking conditions was expected to have profound consequences for pitch perception in binaurally unmasked pure tones. In particular, frequency discrimination in binaurally unmasked tones was expected to be similarly inaccurate as at high frequencies, where temporal fine-structure information is unavailable due to the loss of phase locking (Moore, 1973; see also Siebert, 1970; Heinz et al., 2001a). Moreover, frequency discrimination performance in binaurally unmasked tones was expected to decrease more gradually with increasing stimulus duration than in comparable diotic (homophasic) masking conditions (Siebert, 1970; Heinz et al., 2001a). Finally, frequency discrimination accuracy in binaural unmasking conditions would be expected to be similarly susceptible to disruptions in spectral pitch information as at high frequencies (Moore and Glasberg, 1989; Moore and Sek, 1996). However, the current results did not conform to these expectations; they

showed that, while binaural sluggishness affected the perception of changes in frequency over time, frequency discrimination accuracy, as well as its dependence on stimulus duration and susceptibility to concomitant level changes, was remarkably similar between the homophasic and antiphasic masking conditions.

The current results strongly suggest that pure–tone pitch is processed by the same mechanism in homophasic and antiphasic masking conditions. However, at present, it seems impossible to firmly conclude whether the mechanism is based on temporal or spectral information. Both options would seem to have some plausibility, but also raise some further questions. If pitch is coded temporally in both homophasic and antiphasic masking conditions, one has to assume that, even though binaural sluggishness eliminates pitch-related temporal envelope information (Krumbholz et al., 2009a), it does not affect the fine-structure information. For instance, it is possible that the extraction of fine-structure and envelope information occurs at different stages in the processing hierarchy and that binaural sluggishness succeeds the former and precedes the latter. Alternatively, pitch might be based on spectral cues in both antiphasic and homophasic masking conditions. Spectral cues would be unaffected by binaural sluggishness unless the information changes over time (FM). This would explain why the static frequency discrimination thresholds measured in Experiment 1 (Sec. 4.B.iii) did not differ much between the homophasic and antiphasic conditions, whereas the FM detection thresholds measured in Experiment 2 (Sec. 4.C.iii) were much larger in the antiphasic condition. However, in this account, the difference in frequency discrimination accuracy between low and high frequencies has to be assumed to be caused by factors other than a difference in processing mechanism. For instance, the difference could be due to a

difference in the amount of central processing resources devoted to low and high frequencies. Alternatively, the difference could be due to differences in peripheral processes such as nonlinear compression or the degree to which the cochlear filter shape is asymmetric; at high frequencies, the gain of the cochlear amplifier has been shown to be greater and the resulting compression to span a narrower frequency range than at low frequencies (Lopez-Poveda et al., 2003). These differences could also explain why Moore and Sek (1996) observed an effect of relative modulator phase on MM detection performance at high but not at low frequencies, as well as the unintuitive direction of the effect at high frequencies (see Sec. 4.D.i). The fact that MM detection performance at low frequencies and slow modulation rates is independent of the modulator phase difference can be explained by assuming that FM in these conditions is coded by changes in the peak or centroid of the excitation pattern rather than by changes in excitation level along the flanks of the pattern (Demany and Semal, 1986; Heinz et al., 2001b).

In order to resolve these issues, the next study (**Chapter 5**) explores whether pure–tone pitch at low frequencies is based on temporal or spectral information by measuring frequency discrimination accuracy when one flank of the tone's excitation pattern is obscured by a noise masker.

Chapter 5. Evidence suggesting that the coding of low sound frequencies is based on spectral rather than temporal fine-structure information[5]

5.A INTRODUCTION

In humans, temporal fine-structure information plays a crucial role in binaural hearing, which underpins sound localization and helps to perceive sounds in noisy environments (Licklider, 1948; Wightman and Kistler, 1992; Lavandier and Culling, 2008; for review, see Durlach and Colburn, 1978). Whether temporal fine-structure information plays a similarly important role in monaural processing remains uncertain (see **Chapter 4**). A major problem in investigating this question is that sounds that convey temporal fine-structure information also convey spectral information (Carney and Yin, 1988). Determining whether the auditory system uses the temporal or the spectral information to encode sound frequency is fundamentally important for understanding a wide range of perceptions, particularly pitch, and would be expected to have major implications for the development of pre-processing strategies in assistive hearing devices, such as hearing aids and cochlear implants (Moore and Carlyon, 2005).

As described in **Chapter 4**, the current dominant view is that, at low frequencies, the coding of sound frequency is based on temporal fine-structure rather than spectral information (Moore, 2003, 2008; Plack and Oxenham, 2005). While this would seem beneficial, because the temporal information has been shown to convey frequency with a much higher accuracy than the spectral information (Siebert, 1970; Heinz et al., 2001a), none of the arguments supporting a temporal coding mechanism presented so far seem entirely conclusive. The aim of this study was to re-examine this hypothesis by measuring frequency discrimination accuracy for low-frequency sinusoidal (pure-tone)

[5] Based on Magezi et al. (2009a)

signals, when presented together with filtered noises to mask part of the signals' excitation pattern (Fig. 5.1a). At medium and high sound levels, the excitation pattern of sinusoids becomes asymmetric (black solid line in Fig. 5.1a), with a steeper flank towards the apex of the cochlea (representing lower frequencies) and a shallower flank towards the cochlear base (representing higher frequencies). Either flank can selectively be rendered inaudible through masking with an appropriately-filtered noise: the apical (low-frequency) flank can be masked with a lowpass-filtered noise and the basal (high-frequency) flank can be masked with a highpass-filtered noise (magenta and cyan lines in Fig. 5.1a). If frequency coding at low frequencies were based on temporal fine-structure information, frequency discrimination accuracy would be expected to be independent of the masker type, because the accuracy of the temporal information would not be expected to depend on how the activity is distributed across channels (Moore and Sek, 1996). In contrast, if frequency were coded spectrally, frequency discrimination accuracy would be expected to be significantly better for the highpass than for the lowpass masker, because a small change in frequency would be expected to produce a larger change in excitation level along the steeper apical flank of the signal's excitation pattern, which is left audible by the highpass masker, than the shallower basal flank, which is left audible by the lowpass masker (black solid and dashed lines and arrows in Fig. 5.1a).

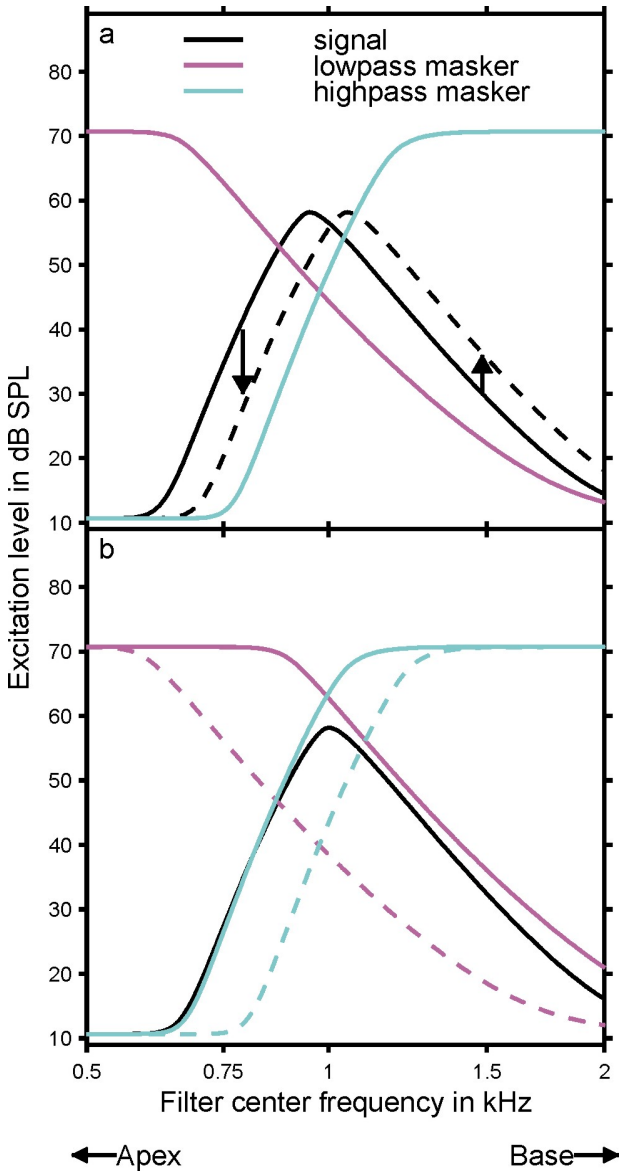

FIG. 5.1 (a) Simulated excitation pattern of a sinusoidal signal at two different frequencies (black solid line: 0.95 kHz; black dashed line: 1.05 kHz), generated using Glasberg and Moore's (1990) level-dependent model of cochlear-filter shape. The arrows show that a given frequency difference creates a larger difference in excitation level along the steeper apical (low-frequency) than the shallower basal (high-frequency) flank of the excitation pattern. The apical or basal flank can be masked (rendered inaudible) by a lowpass- (magenta line) or highpass-filtered noise masker (cyan line), respectively. (b)

When the signal and masker levels are fixed as in the current experiment, the area around the tip of the signal's excitation pattern left audible by the masker can be manipulated by changing the separation between filter cutoff frequency of the masker and the signal frequency (1 kHz). This changes the sensation level (SL) of the signal (solid lines: 5 dB SL; dashed lines: 25 dB SL; sensation levels derived from model simulations rather than from participant data).

5.B EXPERIMENT 1

5.B.i INTRODUCTION

This study consisted of three experiments. In the first experiment, we measured frequency discrimination thresholds for static 500-ms sinusoidal signals around 1 kHz using a standard two-alternative forced-choice task, where participants had to identify the higher-pitched of two successive signals differing in frequency. The signals were presented together with a continuous lowpass- or highpass-filtered noise to mask the apical or basal flank of their excitation pattern, respectively. A crucial parameter in these measurements is the separation between the filter cutoff frequency of the masker and the signal frequency. If the separation were too large (dashed magenta and cyan lines in Fig. 5.1b), the masker would leave too wide a region around the tip of the signal's excitation pattern audible, and no difference in frequency discrimination threshold would be expected between the low- and highpass maskers, even if frequency were in fact coded spectrally. On the other hand, if the separation were too small (solid magenta and cyan lines in Fig. 5.1b), the masker might render one of the two signals in the two-alternative task completely inaudible, and the task would change from a frequency discrimination to a detection task. In this experiment, the separation between the masker cutoff and signal frequency was varied parametrically to find the optimal region in-between these two extremes. The signal and masker levels were set to medium values (overall level of signal: 55 dB SPL; masker spectral density: 60 dB SPL per cochlear-filter bandwidth) to attain a reasonable degree of cochlear-filter asymmetry. With the signal and masker levels fixed in this way, changing the separation between the masker cutoff and signal frequency changes the signal's sensation level (i.e., the difference between the

presentation level of the signal and the level at detection threshold). In order to control for possible detectability cues at low sensation levels, the experiment also contained an "allpass" condition, where the masker was neither low- nor highpass filtered, and the signal sensation level was manipulated by changing the masker spectrum level. In order to be able to determine the signal sensation levels for the masker parameters tested, we also measured the detection threshold of the signal as a function of the masker cutoff frequency and for the allpass masker.

5.B.ii METHODS

Stimuli

In Experiment 1 (current section), all stimuli (signal and masker) were presented diotically (i.e., identically at both ears). The noise maskers used in the current study were filtered so as to produce an approximately constant level of excitation per cochlear-filter bandwidth (measured in terms of the equivalent rectangular bandwidth, or ERB; Glasberg and Moore, 1990) within their passbands. The low- and highpass filters were implemented as "brickwall" filters with cutoff frequencies lower or equal to the signal frequency of 1 kHz in the lowpass conditions, and higher or equal to 1 kHz in the highpass conditions. The filtering was implemented in the frequency domain using a 2^{18}-point fast Fourier transform.

Signal detection thresholds were measured for four different masker cutoff frequencies in both the low- and highpass conditions, with separations of 0, 0.25, 0.5, 1, and 2 ERBs from the 1-kHz signal frequency (grey symbols in Fig. 5.2; the corresponding frequency differences in hertz are shown on the top axis).

Frequency discrimination thresholds were measured for masker cutoff frequencies

with separations of 0, 0.25, 0.5, and 1 ERBs from 1 kHz in the lowpass condition, and 0.25, 0.375, 0.5, and 0.75 ERBs in the highpass condition to yield roughly similar signal sensation levels in both conditions. In the allpass condition, the masker level was 60 dB SPL per ERB for the detection measurements and 55, 52.5, 50, 45 or 40 dB SPL per ERB for the frequency discrimination measurements to cover roughly the same range of signal sensation levels as tested in the low- and highpass conditions. The sensation levels shown on the abscissa of Fig. 5.3 were calculated by linear interpolation of the detection threshold function for the respective masker type (grey symbols in Fig. 5.2).

The signal duration was always 500 ms and included 10-ms squared-cosine ramps. All stimuli were generated digitally at a sampling rate of 25 kHz using TDT System 3 (Tucker-Davies Technology, Alachua, FL) and MATLAB® (The Mathworks, Natick, MA), digital to analogue converted with a 24-bit amplitude resolution (TDT RP2.1), amplified (TDT HB7) and presented over headphones (K240 DF, AKG, Vienna, Austria) in a double-walled sound-attenuating chamber.

Procedure

All threshold measurements used an adaptive two-interval, two-alternate (2I2AFC) procedure with a three-down one-up rule, which tracks 79%-correct performance (Levitt, 1971). The two observation intervals were separated by a 500-ms silent gap. In the detection measurements, only one of the intervals contained the signal (a 1-kHz sinusoid) and the task of the participant was to identify this interval. In these measurements, the adaptive parameter was the signal level. In the frequency discrimination measurements, the adaptive parameter was the frequency difference between the signals, and the task was to identify the interval containing the higher-

frequency signal.

Data Analysis

Many previous studies have converted frequency discrimination or frequency modulation (FM) detection thresholds into logarithmic units, even when the thresholds were expressed in percent as in the current study (e.g., Moore et al., 1996), because threshold variance was found to be heterogeneous on a linear scale, and thus violated the homogeneity-of-variance assumption of the ANOVA (e.g., Howell, 2002). In the current study, frequency discrimination thresholds were evaluated in both linear and logarithmic units, and the degree of variance heterogeneity was analyzed by computing the correlation between the inter-participant means and variances of the thresholds for both (see "Data analysis" in Sec. 4.B.ii). In Experiment 1 (current section), there was a positive correlation for the linear thresholds ($r = 0.540$, $p = 0.070$), and a negative correlation for the logarithmic thresholds ($r = -0.452$, $p = 0.140$), but neither was statistically significant. The ANOVA results were the same for both linear and logarithmic thresholds, and so, only the results for the logarithmic thresholds are presented in the Results section (5.B.iii).

Participants

Six participants (three male, three female, aged between 20 and 34 years) took part in Experiment 1. The participants had no reported history of hearing or neurological disorders. Participants who were not authors of the corresponding manuscript (Magezi et al., 2009a) were paid for their services at an hourly rate. Experimental procedures were approved by the Ethics Committee of Nottingham University School of Psychology.

5.B.iii RESULTS

Frequency discrimination thresholds were generally smaller for the highpass than for the lowpass condition (compare squares and circles in Fig. 5.3). A repeated-measures ANOVA with factors masking condition (lowpass, highpass and allpass) and sensation level (four levels ranging from about 6-19 dB on average) confirmed that the main effect of masking condition was significant [$F(2,10) = 10.193$, $p = 0.004$], and *post hoc* comparisons, using Fisher's least-significant difference test, revealed that this main effect was due to the thresholds for the highpass condition being significantly smaller than those for the lowpass condition ($p = 0.015$). As expected, there was also a main effect of sensation level [$F(3,15) = 68.363$, $p < 0.001$], in that thresholds generally decreased with increasing sensation level. The effect of sensation level was greater in the allpass than the low- or highpass conditions (Fig. 5.3, triangles), as confirmed by a significant interaction between masking condition and sensation level [$F(6,30) = 6.404$, $p < 0.001$]. This was due to the thresholds for the allpass condition increasing sharply for sensation levels around about 5 dB. The fact that a similar increase was not observed for the low- and highpass conditions suggests that participants were using a detectability rather than a pitch cue at these low sensation levels. Finally, there was a trend for the difference between the low- and highpass conditions to decrease with increasing sensation level. This is consistent with the expectation that any difference between the low- and highpass conditions would eventually disappear towards large sensation levels. However, this effect was not entirely consistent across all participants, which is why it was found to be non-significant in an ANOVA of the low- and highpass conditions alone [interaction between masking condition and sensation level $F(3,15) < 0.984$, $p = 0.427$]. As expected,

this ANOVA revealed significant main effects of masking condition [$F(1,5) = 13.272$, $p = 0.015$] and sensation level [$F(3,15) = 24.474$, $p < 0.001$].

The finding that frequency discrimination accuracy was significantly better for the highpass than for the lowpass masker would be predicted if frequency were coded spectrally. However, the difference could also be explained in terms of temporal coding, if one assumes that the auditory system processes temporal fine-structure information by comparing the information across different places along the basilar membrane through a spatial cross-correlation mechanism (Loeb et al., 1983; Shamma, 1985; Deng and Geisler, 1987). This idea is based on the fact that, due to the travelling-wave nature of the basilar-membrane response, the phase of the temporal fine structure of the response changes along the length of the membrane (see **Chapter 1**). As the slope of this phase change is steeper within the apical than the basal part of the response (for review, see Robles and Ruggero, 2001), a spatial cross-correlation mechanism might be expected to yield better frequency discrimination thresholds in the highpass masking condition, which leaves the apical part of the response audible, than in the lowpass condition, which leaves the basal part audible. In order to explore this possibility further, the second experiment investigated how the observed difference in frequency discrimination accuracy between the low- and highpass masking conditions depends on sound level. The asymmetry in the shape of the excitation pattern increases with increasing level (Egan and Hake, 1950). Therefore, if frequency were coded spectrally, an increase in sound level would be expected to lead to a worsening in frequency discrimination accuracy for the lowpass condition and improvement for the highpass condition. In contrast, at high levels the slope of the phase gradient of the travelling-wave response becomes shallower

throughout the apical part, and, on average, steeper in the basal part (Palmer and Shackleton, 2008; see also de Boer and Nuttall; 1997, 2000; Oxenham and Dau, 2001). The result is that at high, but not low, sound levels, the slope of the phase gradient should be similar in both the apical and basal parts of the travelling-wave response. The difference in frequency discrimination accuracy between the low- and highpass conditions would thus be expected to increase with increasing level if frequency were coded spectrally, but to decrease with level if frequency were coded by a spatial cross-correlation mechanism.

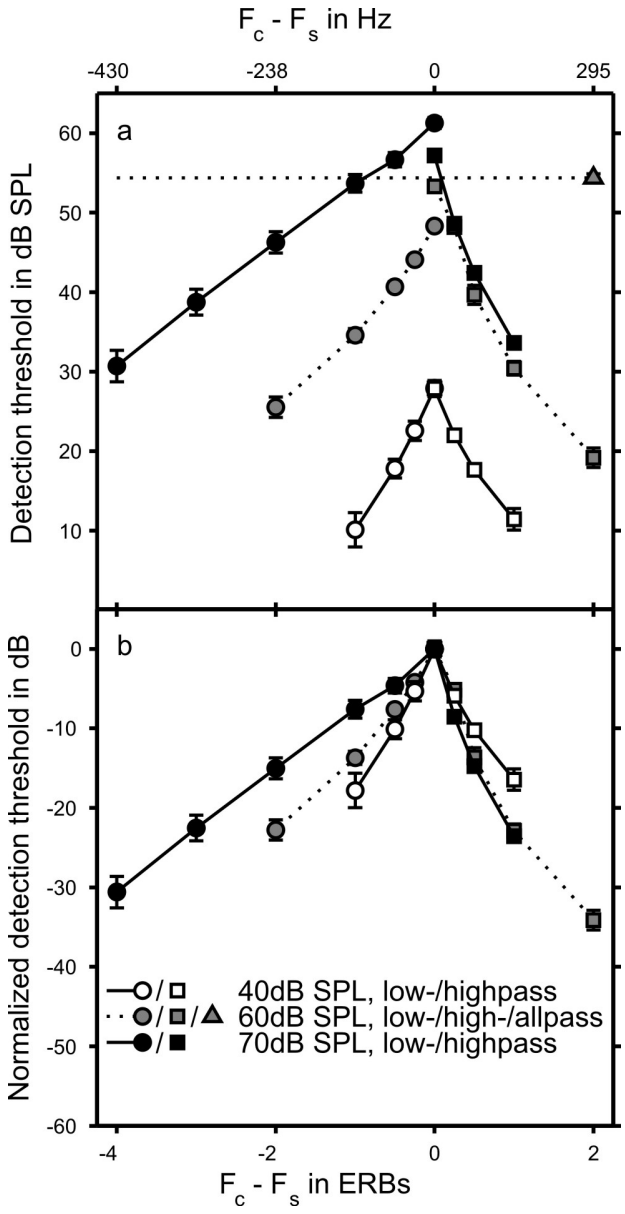

FIG. 5.2 Average detection thresholds for a sinusoidal signal at 1 kHz, masked by a lowpass (circles) or highpass noise masker (squares), plotted as a function of the separation between the masker cutoff (F_c) and signal frequency (F_s = 1 kHz) in number of cochlear-filter bandwidths (equivalent rectangular bandwidths, or ERBs; bottom axis) or hertz (top axis). Different symbol colours and line types show the results for different

masker levels [white and black symbols connected by solid lines: 40 and 70 dB SPL, measured in Experiment 2 (Sec. 5.C); grey symbols connected by dotted lines: 60 dB SPL, measured in Experiment 1 (Sec. 5.B)]. The thresholds are expressed in dB SPL in (a) and normalized to the threshold at zero separation between the masker cutoff and signal frequencies in (b). The threshold for the allpass masker used in Experiment 1 (Sec. 5.B) is shown by the horizontal dotted line and grey triangle.

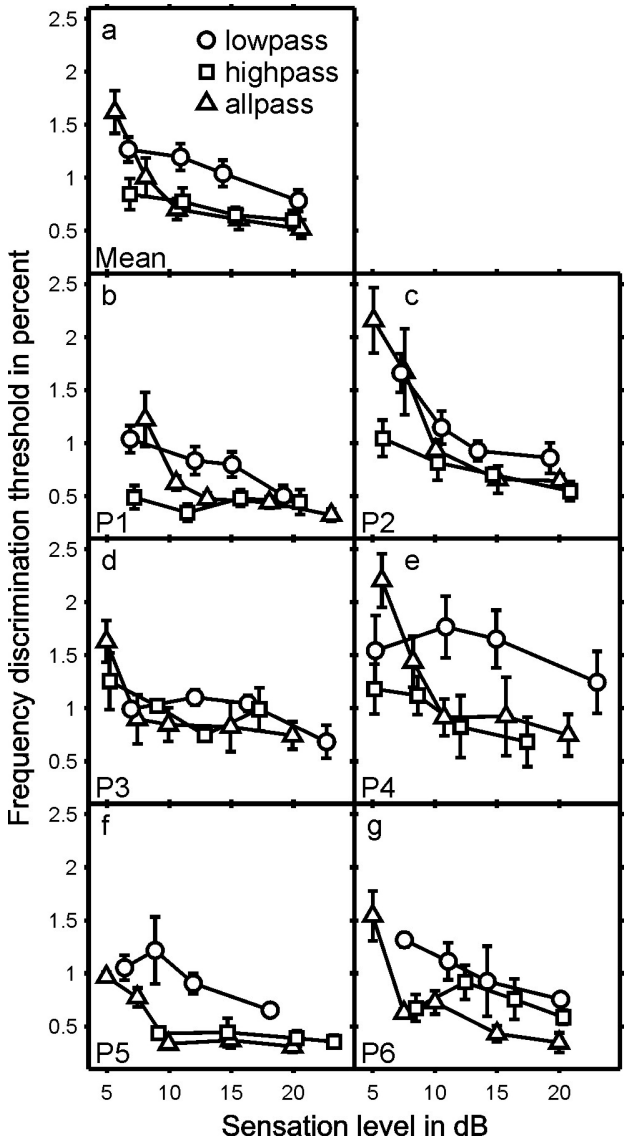

FIG. 5.3 Average (a) and individual (b-e) frequency discrimination thresholds for a 1-kHz sinusoidal signal, masked by a lowpass (circles), highpass (squares) or allpass noise masker (triangles). The thresholds are expressed in percent of the signal frequency and plotted as a function of the signal's sensation level, derived from the detection threshold functions shown in Fig. 5.2. Error bars show the standard error of the mean.

5.C EXPERIMENT 2

5.C.i INTRODUCTION

In Experiment 2, frequency discrimination accuracy was measured with a sinusoidal frequency modulation (FM) detection task, where participants had to identify which of two successive signals was modulated in frequency. As in Experiment 1 (Sec 5.B), the carrier was a 1-kHz sinusoid and was presented in a continuous lowpass- or highpass-filtered noise to mask the basal or apical part of the signal's excitation pattern. The FM was presented at two different rates, a slow rate (2 Hz), where FM is perceived as a change in pitch and is thought to be coded in the same way as the static frequencies used in Experiment 1, and a faster rate (10 Hz), where FM is perceived as roughness or flutter and is thought to be coded spectrally even by advocates of the temporal theory of frequency coding (Moore and Sek, 1994, 1996). Sound level was varied by varying the spectral density of the masker (40 or 70 dB SPL per ERB). The sensation level of the signal was fixed at 10 dB. Based on the results from Experiment 1, a 10-dB sensation level is high enough to avoid the detectability cues that emerge at very low sensation levels, but low enough to yield a sizeable difference in frequency discrimination accuracy between the low- and highpass masking conditions. As in Experiment 1, we first measured the signal's detection threshold as a function of the masker cutoff frequency. In this case, the detection thresholds were used to determine the stimulus parameters (signal level and masker cutoff frequency) to yield a 10-dB sensation level of the signal for the FM detection task.

5.C.ii METHODS

Stimuli

The signal was presented diotically as in Experiment 1, but the masker was uncorrelated between the two ears, because we wanted to use the same masker for a control experiment, in which the signal was presented dichotically (i.e., differently at the two ears). In the detection threshold measurements, four different cutoff frequencies, with separations of 0, 0.25, 0.5, and 1 ERBs from 1 kHz, were used in all conditions, except for the lowpass condition at the higher masker level of 70 dB SPL per ERB, where five cutoff frequencies with larger separations of 0, 0.5, 1, 2, 3, and 4 ERBs from 1 kHz were used to accommodate the shallower slope of the respective threshold function (white and black symbols in Fig. 5.2).

In the FM detection measurements, the signal level was set to the average of the detection thresholds for the low- and highpass maskers with 1-kHz cutoff frequencies (zero separation between the masker cutoff and the signal frequency), and the masker cutoff that would yield a 10-dB signal sensation level for a given condition was determined by linear interpolation of the respective detection threshold function.

The signal duration was 500 ms in the detection measurements, as in Experiment 1, and1000 ms in the FM detection measurements to accommodate the slower modulation rate of 2 Hz. As in Experiment 1, these durations included 10-ms squared-cosine ramps.

The set up and other aspects of the stimulus generation were the same as in Experiment 1 (see Sec 5.B.ii).

Procedure

The detection and FM thresholds were measured with a similar three-down one-up 2I2AFC adaptive procedure as used in Experiment 1. In the FM detection measurements, the adaptive parameter was the FM frequency excursion (difference

between maximal and minimal frequency).

Data Analysis

Frequency modulation detection thresholds were evaluated in both linear and logarithmic units, as in Experiment 1. In Experiment 2, there was a significant correlation between the means and variances for the linear (r = 0.784, p = 0.003), but not the logarithmic thresholds (r = -0.363, p = 0.246), and so, ANOVAs were only performed on the logarithmic thresholds.

Participants

Five participants (two male, three female, aged between 22 and 37 years) took part in Experiment 2, one of whom had also participated in Experiment 1.

5.C.iii RESULTS

Figure 5.2a shows the detection thresholds from Experiment 2 (black and white symbols) as a function of the separation between the masker cutoff (F_c) and signal frequency (F_s = 1 kHz). The detection thresholds from Experiment 1 are also shown for comparison (grey symbols). The threshold patterns, henceforth referred to as "masking patterns", resemble mirrored versions of the signal excitation pattern, with the lowpass-masked thresholds (circles) reflecting the basal (high-frequency) flank of the pattern and the highpass-masked thresholds reflecting the apical (low-frequency) flank (compare Figs 5.1 and 5.2). Normalizing each threshold function to its maximum at zero separation between the masker cutoff and signal frequency (Fig. 5.2b) showed that, like the excitation pattern, the masking pattern became increasingly asymmetric towards higher masker levels (compare white, grey and black symbols in Fig. 3b), in that the lower flank of the pattern became considerably shallower, and the upper flank became slightly

steeper. Determining the slopes of the threshold functions through linear regression and submitting their absolute values to a repeated-measures ANOVA with factors masking condition (low- and highpass) and masker level (40 and 70 dB SPL) showed that this effect was significant [interaction between masking condition and masker level: [$F(1,4) = 96.282$, $p = 0.001$]. Planned comparisons, using *t*-tests, confirmed that the slope of the lower flank of the masking pattern was significantly shallower for the higher (70 dB SPL) than for the lower masker level (40 dB SPL) [$t(4) = 6.196$, $p = 0.002$], whereas that of the upper flank was significantly steeper [$t(4) = -5.264$, $p = 0.006$]. The main effect of masking condition was also significant [$F(1,4) = 14.251$, $p = 0.020$]. Planned comparisons indicated that this effect was due to the upper flank of the masking pattern being steeper than the lower flank for the higher masker level [70 dB SPL; $t(4) = -13.8521$, $p < 0.001$]. However, the same did not apply to the lower masker level [40 dB SPL; $t(4) = 0.595$, $p = 0.584$]. The main effect of masker level was not significant [$F(1,4) = 2.275$, $p = 0.206$].

According to the spectral theory of frequency coding, the FM detection thresholds measured in this experiment should be related to the slope of the audible flank of the signal's excitation pattern (Zwicker, 1970). The results seem to confirm this expectation. At the lower masker level (40 dB SPL), the low- and highpass masking conditions yielded roughly similar FM detection thresholds at both modulation rates tested (2 and 10 Hz; red and blue bars in Fig. 5.4a). As masker level was increased to 70 dB SPL, FM detection thresholds increased in the lowpass condition, and decreased in the highpass condition (red and blue bars in Fig. 5.4b). A repeated-measures ANOVA with factors masking condition (low- and highpass), masker level (40 and 70 dB SPL) and FM rate (2

and 10 Hz), showed that this effect was significant [interaction between masking condition and masker level: $F(1,4) = 27.039$, $p = 0.007$]. Importantly, the level effect on the difference in FM detection performance between the low- and highpass conditions [main effect of masking condition: $F(1,4) = 8.857$, $p = 0.041$], was observed not only for the faster (10 Hz) but also the slower FM rate (2 Hz), as shown by the lack of a significant three-way interaction between masking condition, masker level and FM rate [$F(1,4) = 2.966$, $p = 0.160$]. A correlation analysis showed that, for both FM rates, there was a significant negative correlation of similar magnitude between the FM detection thresholds and the absolute values of the slopes of the relevant detection threshold functions from Fig. 5.2 (Figs 5.5a and 5.5b), as expected based on the spectral theory of frequency coding. The main effect of FM rate was also significant [$F(1,4) = 25.857$, $p = 0.007$]. Neither the main effect of masker level nor the interactions involving FM rate (masker level and FM rate, masking condition and FM rate) were signfificant.

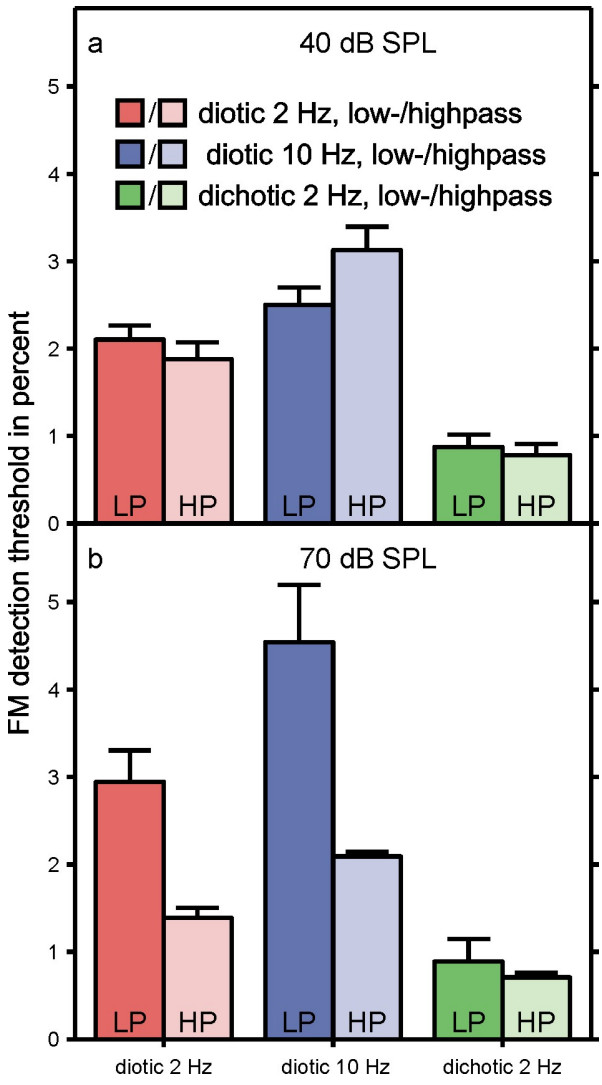

FIG. 5.4 Average FM detection thresholds for a 1-kHz signal and a modulation rate of 2 (red bars) or 10 Hz (blue bars), expressed in terms of the FM frequency excursion in percent of the signal frequency. The left and right (darker- and lighter-shaded) bars in each pair show the results for the lowpass and highpass masking conditions, respectively. The upper panel (a) shows the thresholds for the 40-dB SPL masker and the lower panel (b) shows the thresholds for the 70-dB SPL masker. The green bars in each panel show the results for the dichotic FM detection task [Experiment 3 (Sec. 5.D)], where the modulated signal was replaced by a static signal in one ear.

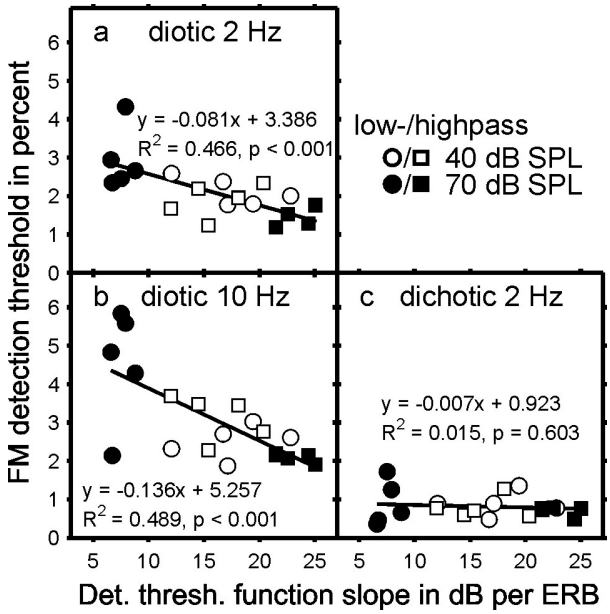

FIG. 5. 5 Linear regression of the individual FM detection thresholds, in percent, for the different conditions and the slopes of the corresponding detection threshold functions, in dB per ERB. Panels (a) and (b) show the correlations for the diotic FM detection task used in Experiment 2 at 2- (a) and 10-Hz (b) modulation rate. Panel (c) shows the correlation for the dichotic control task used in Experiment 3 [2-Hz modulation rate as in (a)]. Different masking conditions and masker levels are shown by different symbols and symbol colours (lowpass: circles; highpass: squares; 40 dB SPL: open symbols; 70 dB SPL: filled symbols).

The results so far suggest that FM detection not only at fast but also at slow modulation rates, as well as static frequency discrimination, are based on spectral rather than temporal fine-structure information. However, this conclusion depends on the assumption that the "goodness" of the temporal information is independent of how the activity is distributed across channels (i.e., the shape of the audible part of the signal's excitation pattern in the current experiments). In order to test this assumption, we conducted a control experiment, which was identical to the 2-Hz FM detection measurements in the current experiment, except that the FM was perceived through (interaural) temporal fine-structure cues.

5.D EXPERIMENT 3

In this experiment, the 2-Hz FM detection measurements from Experiment 2 were repeated with the modulated signal replaced by a static signal in one ear. With the modulated signal presented dichotically in this way, the FM is detected through time-varying interaural temporal cues, which create the perception of motion. These interaural temporal cues are based on the comparison of temporal fine-structure information across the two ears and play a crucial role in human sound localization (see **Chapters 1 and 2**). As in Experiment 2 of the current study, the maskers were uncorrelated between the two ears to avoid confounding the dichotic FM detection thresholds with changes in the detection threshold of the modulated signal caused by binaural unmasking (Hirsch, 1948; see **Chapter 4**). The participants, as well as all other stimulus parameters, were identical to those used in Experiment 2, and the results are shown in the same figures (Figs 5.4 and 5.5).

In agreement with the assumption that the "goodness" of temporal information is indeed independent of the excitation pattern shape, the dichotic FM detection thresholds were little affected by masking condition or masker level (green bars in Fig. 5.4). This was confirmed by a repeated-measures ANOVA of the dichotic thresholds with factors masking condition (low- and highpass) and masker level (40 and 70 dB SPL), which yielded no significant main effects or interaction. Furthermore, there was no significant negative correlation between the dichotic FM detection thresholds and the slopes of the masking pattern flanks as for the diotic thresholds (Fig. 5.5c). In agreement with previous results (Witton et al., 2000), the dichotic FM detection thresholds were considerably smaller than the corresponding diotic thresholds (compare green and red bars in Fig. 5.4). A repeated-measures ANOVA of all FM detection thresholds (diotic and dichotic) with factors masking condition (low- and highpass), masker level (40 and 70 dB SPL), and modulation condition (2- and 10-Hz diotic and 2-Hz dichotic) revealed a significant main effect of modulation condition [$F(2,8) = 49.470$, $p < 0.001$], as well as a significant three-way interaction [$F(2,8) = 9.867$, $p = 0.007$]. Note that the main effects of level and masking condition, and all two-way interactions were not significant.

5.E DISCUSSION

The current results show that frequency discrimination accuracy at low frequencies (1 kHz) is significantly better when based on the steeper apical (low-frequency) than the shallower basal (high-frequency) flank of the excitation pattern. The finding that this difference in frequency discrimination accuracy depends on sound level in a way that is consistent with the level-dependent asymmetry in the slopes of the excitation pattern flanks, and disappears when the frequency information is mediated

through interaural temporal fine-structure cues, means that the difference cannot not be explained in terms of temporal processing. Taken together, the current results provide strong evidence that the coding of low frequencies is based on spectral information, thus conflicting with the current dominant view that that frequency coding at low frequencies is based on temporal fine-structure information (Moore, 2008).

Many of the key arguments in favour of this view do not appear to hold up to closer scrutiny, and it seems that some of the conclusions from previous studies may have been biased by preconception. One of the most-cited arguments is the finding that frequency discrimination accuracy declines sharply above about 4 kHz, and the fact that this decline coincides with what is assumed to be the upper limit of phase locking in humans (Moore, 1973). This argument has two flaws. Firstly, the decline in frequency discrimination accuracy at high frequencies may also be due to a lack of selective pressure for humans to sustain accurate frequency discrimination at high frequencies (Heffner et al., 2001a), because behaviourally relevant sounds like speech and music are mainly composed of low frequencies (Sivian et al., 1959; Byrne et al., 1994). This potential confound makes the differences in frequency discrimination accuracy between low and high frequencies somewhat difficult to interpret. The second problem is that the actual phase locking limit in humans is unknown. The assumed limit of 4 kHz is based on squirrel-monkey data (Rose et al, 1967; Anderson et al., 1971). However, the phase locking limit varies greatly even between mammalian species, ranging from a few hundred hertz in the guinea pig (Palmer and Russell, 1986) to more than 6 kHz in the Jamaican fruit bat (Heffner et al., 2001b). This suggests that the phase locking limit depends crucially on each species' exact ecological requirements. Data on the perception

of fine-structure ITDs suggest that the human phase locking limit may be much lower than 4 kHz. Towards high frequencies, the perception of fine-structure ITDs is limited by two factors, of which the phase locking limit is one, and the other is the ambiguity that arises when the waveform period becomes shorter than twice the maximum possible ITD, determined by head size (Kuhn, 1977). The fact that humans can resolve the head size-related ambiguity in fine-structure ITDs in a similar way as owls, by integrating ITD information across frequencies (Trahiotis and Stern 1989; Saberi et al., 1999), and owls perceive fine-structure ITDs up to their phase locking limit of about 9 kHz (Köppl, 1997), suggests that the human phase locking limit corresponds to the limit of fine-structure ITD perception in humans at about 1.5 rather than 4 kHz (Mossop and Culling, 1998).

Another argument in favour of a dominant role of temporal fine-structure information in frequency coding is the finding that frequency discrimination accuracy at low frequencies is relatively robust to disruptions in spectral frequency cues. Moore and Sek (1996), for instance, measured FM detection thresholds in conditions where the amplitude of the stimulus was also modulated to disrupt spectral FM cues (see Sec. 4.C.i). At slow modulation rates (< 5 Hz), where FM and amplitude modulation (AM) can be tracked as changes in pitch and loudness, respectively, FM detection thresholds were largely unaffected by AM. In contrast, at faster rates (≥ 10 Hz), where both FM and AM elicit a sensation of roughness or flutter, AM caused a substantial degradation in FM detection performance. Moore and Sek explained their results by proposing that FM perception at slow modulation rates is based on temporal fine-structure information, which would be expected to be unaffected by AM. However, the difference in the effectiveness of AM to impair FM detection between slow and fast modulation rates is

more likely to be related to the way in which FM and AM are perceived at different rates, and this may be determined by how they are processed rather than what type of information the processing is based on. For instance, at slow rates, FM may be coded by tracking the peak or centroid of the excitation pattern (Demany & Semal, 1986), which would be similarly unaffected by AM as temporal fine-structure cues (Heinz et al., 2001b). At faster rates, when the changes in pitch or loudness become too fast to track, the perception of both AM and FM may be mediated by modulations in the excitation level of individual frequency channels, as suggested by models of roughness perception (Daniel and Weber, 1997).

Another major challenge in trying to find out which cues the auditory system uses for frequency coding is the inevitable confound between temporal and spectral sound properties through the biunique correspondence of time and frequency represented by the Fourier transform (see, e.g., Zeng et al., 2004). The current study avoided this confound by using masking rather than manipulating the temporal or spectral stimulus properties directly.

The current results suggest that temporal fine-structure information might be mainly used by the binaural system for the analysis of interaural temporal information. This idea is supported by physiological and anatomical data, which have shown that those neurons that best preserve, or even refine, the temporal information conveyed by the auditory nerve, such as the spherical bushy cells in the cochlear nucleus (Oertel, 1983; Joris et al., 1994) or the neurons of the medial nucleus of the trapezoid body (Smith et al., 1998), tend to be found in pathways that are thought to be involved in binaural processing

(Smith et al., 1993). These pathways contain the largest and most temporally accurate synapses in the mammalian brain (Trussell, 1999; Schneggenburger and Forsythe, 2006).

However, the current data do not exclude the possibility that monaural pathways use temporal fine-structure information for processing spectrally complex sounds, such as speech and music, if the processing were inapplicable to spectrally simple sounds as used in the current study. For instance, it has been suggested that pitch perception in complex sounds might be based on a spatial cross-correlation of the temporal responses to different spectral components (Shamma and Klein, 2000). Nevertheless, the current data call for a careful and non-preconceived reconsideration of the role of temporal fine-structure information for monaural processing in both normal and impaired hearing.

General conclusions

The current project comprised five studies investigating the mechanisms by which temporal fine-structure information is processed in the human auditory system. Due to the travelling-wave nature of the cochlear response, the phase of fine-structure information changes along the length of the basilar membrane (for review, see Robles and Ruggero, 2001). The results of the first study (**Chapter 1**) suggest that these cross-channel phase differences may play a crucial role in the processing of ITDs. Participants were able to extract ITDs from pure tones that were partially masked by a highpass noise in one ear and a lowpass noise in the other ear to obscure the basal and apical parts of the tone's cochlear response, respectively. Surprisingly, performance in some participants was strongly asymmetric, depending on which ear received the lowpass noise masker. Model simulations revealed that both the average and the individual data could best be accounted for by a physiological version of Shamma et al.'s (1989) stereausis model, which assumes that ITDs are processed by cross-channel comparisons.

McAlpine and co-workers (2001, 2005; Harper and McAlpine, 2004) suggested that the presence of best ITDs far beyond the physiological range in small-headed mammals could mean that ITDs are coded by an opponent-channel rate code, involving only one channel in each hemisphere, broadly tuned to the contralateral hemispace, rather than many finely tuned channels, as assumed in topographic models (see also von Békésy, 1930; van Bergeijk, 1962). The results of the second study in the current project (**Chapter 2**) suggest that a similar mechanism may also be used in humans. Using the continuous stimulation paradigm (CSP), EEG responses were found to be larger for outward ITD changes than for inward changes. This pattern of results, as well as the

hemispheric distribution of the responses, were highly consistent with the predictions of the opponent-channel model and contravened predictions based on the assumption of a topographic coding of ITDs.

The results from previous fMRI studies would suggest that the ITD change responses measured in the second study arise from a region that lies posterior to primary auditory cortex on planum temporale. This area has been shown to be specifically activated by ITD processing (e.g., Krumbholz, 2005a). The third study in the current project (**Chapter 3**) used the CSP with a rapid event-related fMRI design to show that the topographies of the transient responses to pitch and motion onset overlapped with the topographies of the corresponding sustained responses measured in previous fMRI studies that used conventional epoch-related designs. These results suggest that EEG with the CSP is a valid method to investigate feature coding mechanisms in humans. The results also suggest that the CSP may be a strong candidate paradigm for simultaneous EEG-fMRI recordings.

The final two studies explored the possible role of temporal fine-structure information in the encoding of low sound frequencies using two different approaches. The fourth study (**Chapter 4**) investigated different measures of pure-tone frequency discrimination performance in conditions of binaural unmasking. Based on the finding that binaural sluggishness eliminates temporal envelope cues to pitch in such binaural masking conditions (Krumbholz et al., 2009a), it was expected that binaural sluggishness would also degrade the faster-varying fine-structure cues. However, the results of the fourth study showed little difference in frequency discrimination performance between binaural and diotic masking conditions. These results suggested, either, that binaural

sluggishness does not affect temporal fine-structure cues, or, that frequency coding in pure tones is based on spectral rather than temporal information. The results from the fifth study (**Chapter 5**) suggested that the latter is the case. In this study, frequency discrimination performance for partially masked pure tones was shown to reflect the level-dependent changes in the shape of the pure-tone excitation-pattern. A control experiment showed that processing based on temporal information should have yielded level-independent performance.

The current work has provided new insights on how temporal fine-structure information may be used in binaural processing, which is important for sound localization and listening in noisy environments. At the same time, it has also cast doubt on the common assumption that temporal fine-structure information is used for the coding of frequency in monaural processing, which is important for the perception of pitch, as, for instance, in speech or music. At present there is much research to develop strategies to provide temporal fine-structure information in cochlear implants (e.g. Nie et al., 2005, for review see Moore, 2008), and the results of the current work would seem to have important implications for such developments.

References

Akeroyd M.A., Summerfield A.Q., 1999. A binaural analog of gap detection. J. Acoust. Soc. Am. 105, 2807-2820.

Anderson D. J., Rose J. E., Hind, J. E., Brugge J. F., 1971. Temporal position of discharges in single auditory nerve fibers within the cycle of a sine-wave stimulus: Frequency and intensity effects. J. Acoust. Soc. Am. 49, 1131-1139.

Arnott S.R., Binns M.A., Grady, C.L., Alain C., 2004. Assessing the auditory dual-pathway model in humans. Neuroimage 22, 401-408.

Barrett D.J., Hall, D.A., 2006. Response preferences for "what" and "where" in human non-primary auditory cortex. Neuroimage 32, 968-977.

Batra R., Kuwada S., Fitzpatric D. C., 1997. Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex, I: Heterogeneity of responses. J. Neurophysiol. 78, 1222-1236.

Baumgart F., Gaschler-Markefski B., Woldorff M.G., Heinze H.J., Scheich, H., 1999. A movement-sensitive area in auditory cortex. Nature 400, 724-726.

Bell A.J., Sejnowski T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural. Comput. 7, 1129-1159.

Bernstein J.G., Oxenham A.J., 2003. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? J. Acoust. Soc. Am. 113, 3323-3334.

Bernstein L. R., Trahiotis C., 1996. On the use of the normalized correlation as an index of interaural envelope correlation. J. Acoust. Soc. Am. 100, 1754-1763.

Bernstein L. R., Trahiotis C., 2002. Enhancing sensitivity to interaural delays at

high frequencies by using ‘transposed stimuli’. J. Acoust. Soc. Am. 112, 1026-1036.

Bernstein L. R., Trahiotis C., 2003. Enhancing interaural-delay based extents of laterality at high frequencies by using transposed stimuli. J. Acoust. Soc. Am. 113, 3335-3347.

Blamire A.M., Ogawa S., Ugurbil K., Rothman D., McCarthy, G., Ellermann, J.M., Hyder, F., Rattner, Z., Shulman, R.G., 1992. Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging. Proc. Nat. Acad. Sci. U.S.A. 89, 11069-11073.

Blodgett H.C., Jeffress L.A., Taylor R.W., 1958. Relation of masked threshold to signal-duration for various interaural phase-combinations. Am. J. Psychol. 71,283-290.

Bonham B. H., Lewis E. R., 1999. Localization by interaural time difference ITD: Effects of interaural frequency mismatch. J. Acoust. Soc. Am. 106, 281-290.

Brand A., Behrend O., Marquardt T., McAlpine D., Grothe B., 2002. Precise inhibition is essential for microsecond interaural time difference coding. Nature 417, 543-547.

Breebaart J., van der Par S., Kohlrausch A., 2001. Binaural processing model based on contralateral inhibition. I. Model structure. J. Acoust. Soc. Am. 110, 1074-1088.

Brett M., Johnsrude I.S., Owen A.M., 2002. The problem of functional localization in the human brain. Nat. Rev. Neurosci. 3, 243-249.

Buckner R.L., Bandettini P.A., O'Craven K.M., Savoy, R.L., Petersen S.E., Raichle,

M.E., Rosen B.R., 1996. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. Proc. Nat. Acad. Sci. U.S.A. 93, 14878-14883.

Byrne D., Dillon H., Tran K., Arlinger S., Wilbraham K., Cox R., Hagerman B., Hetu R., Kei J., Lui C., Kiessling J., Kotby M.N., Nasser N.H.A., Elkholy W.A.H., Nakanishi Y., Oyer H., Powell R., Stephens D., Meredith R., Sirimanna T., Tavartkiladze G., Frolenkov G.I., Westerman S., Ludvigsen C., 1994. An international comparison of long-term average speech spectra. J. Acoust. Soc. Am. 96, 2108-2120.

Carlyon R. P., Shamma S., 2003. An account of monaural phase sensitivity. J. Acoust. Soc. Am. 114, 333-348.

Carney L. H., Yin T. C., 1988. Temporal coding of resonances by low-frequency auditory nerve fibers: Single-fibre responses and a population model, J. Neurophysiol. 60, 1653-1677.

Carr C.E., Konishi M., 1990. A circuit for detection of interaural time differences in the brain stem of the barn owl. J. Neurosci. 10, 3227-3246.

Chait M., Poeppel, D., de Cheveigné, A., Simon, J.Z., 2005. Human auditory cortical processing of changes in interaural correlation. J. Neurosci. 25, 8518- 8527.

Chait M., Poeppel D., de Cheveigné A., Simon J.Z., 2007. Processing asymmetry of transactions between order and disorder in human auditory cortex. J. Neurosci. 27, 5207-5214

Chait M., Poeppel D., Simon J.Z., 2008. Auditory temporal edge detection in human auditory cortex. Brain Res. 1213, 78-90.

Clarke S., Bellmann A., Meuli R.A., Assal G., Steck A.J., 2000 Auditory agnosia and auditory spatial deficits following left hemispheric lesions: Evidence for distinct processing pathways. Neuropsychologia 38, 797-807.

Colburn H.S., 1973. Theory of binaural interaction based on auditory-nerve data. I. General strategy and preliminary results on interaural discrimination. J. Acoust. Soc. Am. 54, 1458-1470.

Colburn H. S., 1996. Computational models of binaural processing, in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Poper, R. R. Fay. Springer, New York, pp. 332-400.

Colburn H.S., Latimer S.J., 1978. Theory of binaural interaction based on auditory-nerve data. III. Joint dependence on interaural time and amplitude differences in discrimination and detection. J. Acoust. Soc. Am 64, 95-106.

Culling J.F., Colburn H.S., 2000. Binaural sluggishness in the perception of tone sequences and speech in noise. J. Acoust. Soc. Am. 107, 517-527.

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

Dai H., 2000. On the relative influence of individual harmonics on pitch judgment. J. Acoust. Soc. Am. 107, 953-959.

Dale A.M., Buckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. Hum. Brain Mapp. 5, 329-340.

Daniel P., Weber R., 1997. Psychoacoustical roughness: Implementation of an optimized model. Acustica 83, 113-123.

de Boer E., Nuttall A.L., 1997. The mechanical waveform of the basilar membrane. I. Frequency modulations ('glides') in impulse responses and cross-correlation functions. J. Acoust. Soc. Am. 101, 3583-3592.

de Boer E., Nuttall A.L., 2000. The mechanical waveform of the basilar membrane. III. Intensity effects. J. Acoust. Soc. Am. 107, 1497-1507.

Deatherage B. H., 1961. Binaural interaction of clicks of different frequency content. J. Acoust. Soc. Am. 33, 139-145.

Deatherage, B. H., 1966. Examination of binaural interaction. J. Acoust. Soc. Am. 39, 232-249.

Delorme A., Makeig S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9-21.

Demany L., Semal C., 1986. On the detection of amplitude modulation and frequency modulation at low modulation frequencies. Acustica 61, 243-255.

Deng L. & Geisler C.D., 1987. A composite auditory model for processing speech sounds. J. Acoust. Soc. Am. 82, 2001-2012.

Deouell L.Y., Bentin S., Giard M.H., 1998. Mismatch negativity in dichotic listening: Evidence for interhemispheric differences and multiple generators. Psychophysiology 35, 355-365.

Deouell L.Y., Heller A.S., Malach, R., D'Esposito, M., Knight, R.T., 2007. Cerebral responses to change in spatial location of unattended sounds. Neuron 55, 985-996.

Doeller C.F., Opitz B., Mecklinger A., Krick C., Reith W., Schröger E., 2003.

Prefrontal cortex involvement in preattentive auditory deviance detection: Neuroimaging and electrophysiological evidence. Neuroimage 20, 1270-1282.

Domnitz R. H., Colburn H. S., 1977. Lateral position and interaural discrimination. J. Acoust. Soc. Am. 61, 1586-1598.

Durlach N.I., 1972. Equilization and cancellation theory, in *Founations of modern auditory theory*, Vol II, edited by J.V. Tobias. Academic, New York, pp. 371-462.

Durlach N.I., Colburn H.S., 1978. Binaural phenomena, in *Handbook of perception* edited by E.C. Carterette and M.P. Friedman. Academic Press, New York, pp. 405-466.

Egan J. P., Hake H. W., 1950. On the masking pattern of a simple auditory stimulus. J. Acoust. Soc. Am. 22, 622-630.

Eggermont J.J., 2002 Temporal modulation transfer functions in cat primary auditory cortex: Separating stimulus effects from neural mechanisms. J. Neurophysiol. 87, 305-321.

Eickhoff S.B., Stephan K.E., Mohlberg H., Grefkes C., Fink G.R., Amunts K., Zilles K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25, 1325-1335.

Fitzpatrick D.C., Kuwada S., 2001. Tuning to interaural time differences across frequency. J. Neurosci 21, 4844-4851.

Fitzpatrick D. C., Kuwada, S., Batra, R., 2000. Neural sensitivity to interaural time differences: Beyond the Jeffress model. J. Neurosci. 20, 1605-1615.

Fitzpatrick D.C., Kuwada S., Batra R., 2002. Transformations in processing interaural

time differences between the superior olivary complex and inferior colliculus: Beyond the Jeffress model. Hear Res 168, 79-89.

Frahm J., Dechent P., Baudewig K., Merboldt, K.D., 2004. Advances in functional MRI of the human brain. Prog. NMR Spect. 44, 1-32.

Freyman RL, Nelson DA. Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners. J. Acoust. Soc. Am. 79:1034-1044, 1986

Friauf E., Lohmann C., 1999. Development of auditory brainstem circuitry. Activity-dependent and activity-independent processes. Cell Tissue Res. 297, 187-195.

Friston K.J., Fletcher P., Josephs O., Holmes A., Rugg M.D., Turner R., 1998. Event-related fMRI: Characterizing differential responses. Neuroimage 7, 30-40.

Friston K.J., Holmes, A.P., Worsley K.J., Poline J.B., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189-210.

Friston K.J., Zarahn E., Josephs, O., Henson R.N., Dale A.M., 1999. Stochastic designs in event-related fMRI. Neuroimage 10, 607-619.

Gaab N., Gabrieli J.D., Glover G.H., 2007a. Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Hum. Brain Mapp. 28, 703-720.

Gaab N., Gabrieli J.D., Glover G.H., 2007b. Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory

processing in the absence and presence of recorded scanner noise using a sparse design. Hum. Brain Mapp. 28, 721-732.

Gebhardt C.J., Goldstein D.P., Robertson R.M., 1972. Frequency discrimination and the MLD. J. Acoust. Soc. Am. 51, 1228-1232.

Glasberg B. R., Moore B. C., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103-138.

Grantham D.W., Wightman F.L., 1979. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. J. Acoust. Soc. Am. 65, 1509-1517.

Green D.M., 1966. Interaural phase effects in masking of signals of different durations. J. Acoust. Soc. Am. 39,720-724.

Green D.M., 1973. Minimum integration time, in *Basic mechanisms in hearing*, edited by A.R. Møller, P. Boston. Academic Press, New York, pp. 829 - 846.

Greenwood D. D., 1971. Aural combination tones and auditory masking. J. Acoust. Soc. Am. 50, 502-543.

Greenwood D. D., 1990. A cochlear frequency-position function for several species: 29 years later. J. Acoust. Soc. Am. 87, 2592-2605.

Griffin S.J., Bernstein L.R., Ingham N.J., McAlpine D., 2005. Neural sensitivity to interaural envelope delays in the inferior colliculus of the guinea pig. J. Neurophys. 93, 346-3478.

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

Grothe B., 2003. New roles for synaptic inhibition in sound localization. Nat. Rev.

Neurosci. 4, 540-550.

Gutschalk A., Patterson R.D., Scherg M., Uppenkamp S., Rupp A., 2004. Temporal dynamics of pitch in human auditory cortex. Neuroimage 22, 755-766.

Hall D.A., Haggard M.P., Akeroyd M.A., Palmer A.R., Summerfield A.Q., Elliott M.R., Gurney E.M., Bowtell R.W., 1999. 'Sparse' temporal sampling in auditory fMRI. Hum. Brain. Mapp. 7, 213-223.

Hall D.A., Hart H.C., Johnsrude I.S., 2003. Relationships between human auditory cortical structure and function. Audiol. Neurootol. 8, 1-18.

Hall D.A., Plack C.J., 2009. Pitch Processing Sites in the Human Auditory Brain. Cereb. Cortex 19, 576-585.

Hall J.W., Grose J.H., 1992. Masking release for gap detection. Philos. Trans. R. Soc. Lond., Ser B 336, 331-337.

Hall J.W., Grose J.H., Dev M.B., 1997. Signal detection and pitch ranking in conditions of masking release. J. Acoust. Soc. Am. 102,1746-1754.

Hall J.W., Wood E.J., 1984. Stimulus duration and frequency discrimination for normal-hearing and hearing-impaired subjects. J. Speech Hear. Res. 27, 252-256.

Halliday, R., Callaway, E., 1978. Time shift evoked potentials (TSEPs): Method and basic results. Electroencephalogr. Clin. Neurophysiol. 45, 118-121.

Harms M.P., Guinan J.J., Jr., Sigalovsky I.S., Melcher J.R., 2005. Short-term sound temporal envelope characteristics determine multisecond time patterns of activity in human auditory cortex as shown by fMRI. J. Neurophysiol. 93, 210-222.

Harms M.P., Melcher J.R., 2002. Sound repetition rate in the human auditory pathway:

Representations in the waveshape and amplitude of fMRI activation. J. Neurophysiol. 88, 1433-1450.

Harms M.P., Melcher J.R., 2003. Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Hum. Brain Mapp. 20, 168-183.

Harper N.S., McAlpine D., 2004. Optimal neural population coding of an auditory spatial cue. Nature 430, 682-686.

Hausmann M., Corballis M.C., Fabri M., Paggi A., Lewald J., 2005. Sound lateralization in subjects with callostomy, callosal agenesis, or hemispherectomy. Brain Res. Cogn. Brain Res. 25, 537-546.

Heinz M.G., Colburn H.S., Carney L.H., 2001a. Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. Neural. Comput. 13, 2273-2316.

Heinz M.G., Colburn H.S., Carney L.H., 2001b. Evaluating auditory performance limits: II. One-parameter discrimination with random-level variation. Neural. Comput. 13, 2317-2338.

Heffner R.S., Koay G., Heffner H.E., 2001a. Audiograms of five species of rodents: implications for the evolution of hearing and the perception of pitch. Hear. Res. 157, 138-152.

Heffner R.S., Koay G., Heffner H.E., 2001b. Sound localization in a new-world frugivorous bat, Artibeus jamaicensis: Acuity, use of binaural cues, and relationship to vision. J. Acoust. Soc. Am. 109, 412-421.

Helmholtz H.L.F., 1863. Die Lehre von den Tonemfindugen als physiologische Grundlage für die Theorie der Musik. Vieweg, Braunschweig (Translated by Ellis A.J., 1954. On the sensations of tone. Dover, New York.)

Henning G.B., 1970. A comparison of the effects of signal duration on frequency and amplitude discrimination, in *Frequency analysis and periodicity in hearing* edited by R. Plomp, and G.F. Smoorenburg. A.W. Sijthoff, Leiden, pp. 350 - 361.

Henning G.B., 1990. The effect of interaural phase on frequency discrimination with broad- and narrow-band maskers. Hear. Res. 48, 195-200.

Henning G.B., Wartini S., 1990. The effect of signal duration on frequency discrimination at low signal-to-noise ratios in different conditions of interaural phase. Hear. Res. 48, 201-207.

Herrmann C.S., Debener, S., 2008. Simultaneous recording of EEG and BOLD responses: A historical perspective. Int. J. Psychophysiol. 67, 161-168.

Hewson-Stoate N., Schönwiesner M., Krumbholz K., 2006. Vowel processing evokes a large sustained response anterior to primary auditory cortex. Eur. J. Neurosci. 24, 2661-2671.

Hine, J., Debener, S., 2007. Late auditory evoked potentials asymmetry revisited. Clin. Neurophysiol. 118, 1274-1285.

Hirsh I.J., 1948. The influence of interaural phase on interaural summation. J. Acoust. Soc. Am. 20, 536-544

Houtgast, T., 1972. Psychophysical evidence for lateral inhibition in hearing. J. Acoust. Soc. Am. 51, 1885-1894.

Houtgast, T., 1973. Psychophysical experiments on ‘tuning curves’ and ‘two-tone

inhibition'. Acustica 29, 168-179.

Houtsma, A.J.M., Smurzynski, J., 1990. Pitch identification and discrimination for complex tones with many harmonics. J. Acoust. Soc. Am. 87, 304-310.

Howell D.C., 2002. Statistical methods for psychology. Thompson, Pacific Grove, CA.

Jäncke, L., Wüstenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage 15, 733-746.

Jeffress L. A., 1948. A place theory of sound localization. J. Comp. Physiol. Psychol. 41, 35-39.

Johnson D.H., 1980. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J. Acoust. Soc. Am. 68, 1115-1122.

Jones S.J., Pitman J.R., Halliday A.M., 1991. Scalp potentials following sudden coherence and discoherence of binaural noise and change in the inter-aural time differences: A specific binaural evoked potential or a 'mismatch' response? Electorencephaolgr. Clin. Neurophysiol. 80, 146-154.

Joris P.X., Carney L.H., Smith P.H., Yin T.C., 1994. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. J. Neurophysiol. 71, 1022-1036.

Joris P. X., Van de Sande B., Louage D. H., van der Heijden M., 2006. Binaural and cochlear disparities. Proc. Natl. Acad. Sci. U.S.A. 103, 12917-12922.

Joris P. X., van der Heijen M., Louage, D., Van der Sande B., Van Kerckhoven C., 2004. Dependence of binaural and cochlear 'best delays' on characteristic

frequency, in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet. Springer, New York, pp. 396-402.

Joris P. X., Yin T. C., 1995. Envelope coding in the lateral superior olive, I: Sensitivity to interaural time differences. J. Neurophysiol. 73, 1043-1062.

Joris P.X., Yin T. C., 2007. A matter of time: Internal delays in binaural processing. Trends Neurosci. 30, 70-78.

Kaas J.H., Hackett, T.A., 2000. Subdivisions of auditory cortex and processing streams in primates. Proc. Nat. Acad. Sci. U.S.A. 97, 11793-11799.

Kaiser J., Lutzenberger W., Preissl H., Ackermann H., Birbaumer N., 2000. Right-hemisphere dominance for the processing of sound-source lateralization. J. Neurosci. 20, 6631-6639.

Klumpp R.G., Eady H.R., 1956. Some measurements of interaural time difference thresholds. J. Acoust. Soc. Am. 28, 859-860.

Knudsen E.I., du Lac S., Esterly S.D., 1987. Computational maps in the brain. Annu. Rev. Neurosci. 10, 45-65.

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

König E., 1957. Effect of time on pitch discrimination thresholds under several psychophysical procedures - comparison with intensity discrimination thresholds. J. Acoust. Soc. Am. 29, 606-612.

Köppl C., 1997. Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, Tyto alba. J. Neurosci. 17, 3312-3321.

Konishi M., 2003. Coding of auditory space. Annu. Rev. Neurosci. 26, 31-55.

Krumbholz K., Hewson-Stoate N., Schönwiesner M., 2007. Cortical response to auditory motion suggests an asymmetry in the reliance on inter-hemispheric connections between the left and right auditory cortices. J. Neurophysiol. 97, 1649-1655.

Krumbholz K., Magezi D.A., Moore R.C., Patterson R.D., 2009a. Binaural sluggishness precludes pitch processing based on envelope cues in conditions of binaural unmasking. J. Acoust. Soc. Am. 125, 1067-1074.

Krumbholz K., Nobis E.A., Weatheritt R.J., Fink G.R., 2009b. Executive control of spatial attention shifts in the auditory compared to the visual modality. Hum. Brain Mapp. 30, 1457-1469.

Krumbholz K., Patterson R.D., Pressnitzer D., 2000. The lower limit of pitch as determined by rate discrimination. J. Acoust. Soc. Am. 108, 1170-1180.

Krumbholz K., Patterson R.D., Seither-Preisler A., Lammertmann C., Lutkenhoner B., 2003. Neuromagnetic evidence for a pitch processing centre in Heschl's gyrus. Cereb. Cortex 13, 765-772.

Krumbholz K., Schönwiesner M., von Cramon D.Y., Rübsamen R., Shah N.J., Zilles K., Fink G.R., 2005a. Representation of interaural temporal information from left and right auditory space in the human planum temporal and inferior parietal lobe. Cereb. Cortex 15, 317-324.

Krumbholz K., Schonwiesner M., Rubsamen R., Zilles K., Fink G.R., von Cramon D.Y., 2005b. Hierarchical processing of sound location and motion in the human brainstem and planum temporale. Eur J. Neurosci. 21, 230-238.

Kuhn G.F., 1977. Model for interaural time differences in the azimuthal plane. J. Acoust. Soc. Am. 62, 157-167.

Lavandier M., Culling J.F., 2008. Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer. J. Acoust. Soc. Am. 123, 2237-2248

Lee T., Girolami M., Sejnowski T., 1999. Independent component analysis using an extended infomax algorithm for mixed subgaussian sources. Neural Comput. 11, 417-441.

Leonard C.M., Puranik C., Kuldau, J.M., Lombardino L.J., 1998. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? Cereb. Cortex 8, 397-406.

Levitt H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467-477.

Liang C., Chistovich L.A., 1961. Frequency-difference limens as a function of tonal duration. Sov. Phys. Acoust. 6, 75-80.

Licklider J.C.R., 1948. The influence of interaural phase relations upon the masking of speech by white noise. J. Acoust. Soc. Am. 20, 150-159.

Lindemann W., 1986. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. J. Acoust. Soc. Am. 80, 1608-1622.

Liu T.T., Frank L.R., Wong E.C., Buxton R.B., 2001. Detection power, estimation efficiency, and predictability in event-related fMRI. Neuroimage 13, 759-773.

Loeb G.E., White M.W., & Merzenich M.M., 1983. Spatial cross-correlation. A proposed mechanism for acoustic pitch perception. *Biol. Cybern* 47, 149-163.

Logothetis N.K., Pauls J., Augath M., Trinath T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150-157.

Logothetis N.K., Wandell B.A., 2004.Interpreting the BOLD signal. Annu. Rev. Physiol. 66, 735-769.

Lomber S.G., Malhotra, S., Hall, A.J., 2007. Functional specialization in non-primary auditory cortex of the cat: Areal and laminar contributions to sound localization. Hear. Res. 229, 31-45.

Lopez-Poveda E.A., Plack C.J., Meddis R., 2003. Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing. J. Acoust. Soc. Am. 113, 951-960.

McAlpine D., 2005. Creating a sense of auditory space. J. Physiol. 566, 21-28.

McAlpine D., Grothe B., 2003. Sound localization and delay lines: Do mammals fit the model? Trends Neurosci. 26, 347-350.

McAlpine D., Jiang D., Palmer A. R., 1996. Interaural delay sensitivity and the classification of low best-frequency binaural responses in the inferior colliculus of the guinea pig. Hear. Res. 97, 136-152.

McAlpine D., Jiang, D., Palmer, A. R., 2001. A neural code for low-frequency sound localization in mammals. Nat. Neurosci. 4, 396-401.

Magezi D.A., Krumbholz K., 2008. Can the binaural system extract fine-structure interaural time differences from noncorresponding frequency channels? J. Acoust. Soc. Am. 124, 3095-3107.

Magezi D.A., Krumbholz K., 2009a. A new paradigm for measuring feature-specific auditory cortical responses with rapid event-related fMRI. (submitted)

Magezi D.A., Krumbholz K., 2009b. Evidence for opponent-channel coding of interaural temporal cues to sound lateralisation in human auditory cortex. (submitted)

Magezi D.A., BoSmith I., Krumbholz K., 2009a. Evidence suggesting that the coding of low sound frequencies is based on spectral rather than temporal fine-structure information. (submitted)

Magezi D.A., Moore R.C, Ponting S.H.Z., Krumbholz K., 2009b. Does binaural sluggishness affect processing in binaurally unmasked low-frequency pure tones? (submitted)

Malhotra S., Hall A.J., Lomber S.G., 2004. Cortical control of sound localization in the cat: Unilateral cooling deactivation of 19 cerebral areas. J. Neurophysiol. 92, 1625-1643.

Malhotra S., Stecker G.C., Middlebrooks J.C., Lomber S.G., Sound localization deficits during reversible deactivation of primary auditory cortex and/or the dorsal zone. J. Neurophysiol. 99, 1628-1642.

Martin B.A., Boothroyd A., 1999. Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. Ear Hear. 20, 33-44.

Martin B.A., Boothroyd A., 2000. Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. J. Acoust. Soc. Am. 107, 2155-2161.

Micheyl C., Moore B.C.J., Carlyon R.P., 1998. The role of excitation-pattern cues

and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones. J. Acoust. Soc. Am. 104, 1039-1050.

Molholm S., Martinez A., Ritter W., Javitt D.C., Foxe, J.J., 2005. The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. Cereb. Cortex 15, 545-551.

Moore B.C.J. , 1973. Frequency difference limens for short-duration tones. J. Acoust. Soc. Am. 54, 610-619.

Moore B.C.J., 2003. An introduction to the psychology of hearing. Academic Press, London.

Moore B.C.J., 2008. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. J. Assoc. Res. Otolaryngol. 9, 399-406.

Moore B.C.J. & Carlyon R.P., 2005. Perception of pitch by people with cochlear hearing loss and by cochlea implant users in *Pitch: Neural Coding and Perception,* edited by C.J. Plack, A.J. Oxenham, and R.R. Fay. Springer, New York, pp 234 - 270.

Moore B.C.J., Glasberg B.R., 1989. Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation. J. Acoust. Soc. Am. 86, 1722-1732.

Moore B.C.J., Sek A., 1992. Detection of combined frequency and amplitude modulation. J. Acoust. Soc. Am. 92, 3119-3131.

Moore B.C.J., Sek A., 1994. Effects of carrier frequency and background noise on the detection of mixed modulation. J. Acoust. Soc. Am. 96, 741-751.

Moore B.C.J., Sek A., 1996. Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. J. Acoust. Soc. Am. 100, 2320-2331.

Moore B.C.J., Sek A., 1998. Discrimination of frequency glides with superimposed random glides in level. J. Acoust. Soc. Am. 104, 411-421.

Morosan P., Rademacher J., Schleicher A., Amunts K., Schormann T, Zilles K., 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage 13, 684-701.

Mossop J. E., Culling J. F., 1998. Lateralization of large interaural delays. J. Acoust. Soc. Am. 104, 1574-1579.

Näätänen R., Gaillard A.W., Mäntysalo S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. (Amst) 42, 313-329.

Näätänen R., Winkler I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. Psychol. Bull. 125, 826-859

Nebel K., Stude P., Wiese H., Müller B., de Greiff A., Forsting M., Diener H.C., Keidel M., 2005. Sparse imaging and continuous event-related fMRI in the visual domain: A systematic comparison. Hum. Brain Mapp. 24, 130-143.

Nichols T., Brett M., Andersson J., Wager T., Poline J.B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653-660.

Nie K., Stickney G., Zeng F.G., 2005. Encoding frequency modulation to improve cochlear implant performance in noise. IEEE Trans. Biomed. Eng. 52, 64-73.

Novitski N., Alho K., Korzyukov O., Carlson S., Martinkauppi S., Escera C., Rinne

T., Aronen H.J., Naatanen, R., 2001. Effects of acoustic gradient noise from functional magnetic resonance imaging on auditory processing as reflected by event-related brain potentials. Neuroimage 14, 244-251.

Novitski N., Maess B., Tervaniemi M., 2006. Frequency specific impairment of automatic pitch change detection by fMRI acoustic noise: An MEG study. J. Neurosci. Methods 155, 149-159.

Nuetzel J.M., Hafter E.R., 1976. Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration. J. Acoust. Soc. Am. 60, 1339-1346.

Nuttall A. L., Dolan D. F., 1993. Two-tone suppression of inner hair cell and basilar membrane responses in the guinea pig. J. Acoust. Soc. Am. 93, 390-400.

O'Mard L. P., Meddis R., 2004. A new visual dimension to auditory modelling using DSAM. Assoc. Res. Otolaryngol. Abstr. 27, 312-313.

Oertel D., 1983. Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus. J. Neurosci. 3, 2043-2053.

Oertel D., 1997. Encoding of timing in the brain stem auditory nuclei of vertebrates. Neuron 19, 959-962.

Oertel D., 1999. The role of timing in the brain stem auditory nuclei of vertebrates, Annu. Rev. Physiol. 61, 497-591.

Oldfield R. C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97-113.

Opitz B., Mecklinger A., Friederici A.D., von Cramon D.Y., 1999a. The functional neuroanatomy of novelty processing: Integrating ERP and fMRI results. Cereb. Cortex 9, 379-391.

Opitz B., Mecklinger A., Von Cramon D.Y., Kruggel F., 1999b. Combining electrophysiological and hemodynamic measures of the auditory oddball. Psychophysiology 36, 142-147.

Opitz B., Rinne T., Mecklinger A., von Cramon D.Y., Schroger E., 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. Neuroimage 15, 167-174.

Overholt E.M., Rubel E.W., Hyson R.L., 1992. A circuit for coding interaural time differences in the chick brainstem. J. Neurosci. 12, 1698-1708.

Oxenham A.J. & Dau T., 2001. Towards a measure of auditory-filter phase response. J. Acoust. Soc. Am. 110, 3169-3178.

Palmer A.R., Bullock D.C., Chambers J.D., 1998. A high-output, high-quality sound system for use in auditory fMRI. Neuroimage 7, S359.

Palmer A.R., Russell I.J., 1986. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hear. Res. 24, 1-15.

Palmer A.R., Shackleton T.M., 2008. Variation in the phase of response to low-frequency pure tones in the guinea pig auditory nerve as functions of stimulus level and frequency. J. Assoc. Res. Otolaryngol. 10, 233-250.

Patterson R. D., 1994. The sound of a sinusoid: Spectral models. J. Acoust. Soc. Am. 96, 1409-1418.

Patterson R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron 36, 767-776.

Penagos H., Melcher J.R., Oxenham A.J., 2004. A neural representation of pitch

salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J. Neurosci. 24, 6810-6815.

Plack C.J., Oxenham A.J., 2005. The psychophysics of pitch, in *Pitch : neural coding and perception* edited by C.J. Plack, A.J. Oxenham, R.R. Fay, and A.N. Popper. Springer, New York, pp. 7-55.

Plomp R., 1965. Detectability threshold for combination tones. J. Acoust. Soc. Am. 37, 1110-1123.

Pressnitzer D., Patterson R.D., Krumbholz K., 2001. The lower limit of melodic pitch. J. Acoust. Soc. Am. 109, 2074-2084.

Rauschecker J.P., Tian B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Nat. Acad. Sci. U.S.A. 97, 11800-11806.

Recanzone G.H., 2000. Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. Hear. Res. 150, 104-118.

Ren T., 2002.Longitudinal pattern of basilar membrane vibration in the sensitive cochlea. Proc. Natl. Acad. Sci. U.S.A. 99, 17101-17106.

Ritsma R.J., 1967. Frequencies dominant in perception of pitch of complex sounds. J. Acoust. Soc. Am. 42, 191-198.

Robinson D.E., Trahiotis C., 1972. Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions. Percept. Pscychophys. 12, 333-334.

Robles L., Ruggero M. A., 2001. Mechanics of the mammalian cochlea. Physiol. Rev. 81, 1305-1352.

Rose J.E., Brugge J.F., Anderson D.J., & Hind J.E., 1967. Phase-locked response to low-frequency tones in single auditory nerve fibres of the squirrel monkey. J. Neurophysiol. 30, 769 -793.

Ruggero M. A., Rich N. C., Recio A., Narayan S. S., Robles, L. 1997. Basilar-membrane responses to tones at the base of the chinchilla cochlea. J. Acoust. Soc. Am. 101, 2151-2163.

Saberi K., Takahashi Y., Farahbod H., & Konishi M., 1999. Neural bases of an auditory illusion and its elimination in owls. Nat. Neurosci. 2, 656-659.

Scharf B., Florentine M., Meiselman C. H., 1976. Critical band in auditory lateralization. Sens. Processes 1, 109-126.

Scherg M., Picton T.W., 1991. Separation and identification of event-related potential components by brain electric source analysis. Electroencephalogr. Clin. Neurophysiol. Suppl. 42, 24-37.

Schneggenburger R., Forsythe I.D., 2006. The calyx of Held. Cell Tissue Res 326, 311-337.

Schönwiesner M., Krumbholz K., Rubsamen R., Fink G.R., von Cramon D.Y., 2007a. Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. Cereb. Cortex 17, 492-499.

Schönwiesner M., Novitski N., Pakarinen S., Carlson S., Tervaniemi M., Naatanen R., 2007b. Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. J. Neurophysiol. 97, 2075-2082.

Schroeder M. R., 1977. New viewpoints in binaural interactions, in *Psychophysics and*

Physiology of Hearing, edited by E. F. Evans and J. P. Wilson. Academic, New York, pp. 455-467.

Schubert E. D., Elpern B. S., 1959. Psychophysical estimate of the velocity of the travelling wave. J. Acoust. Soc. Am. 31, 990-994.

Seebeck A., 1843. Ueber die Sirene. Ann. Phys. 136, 449-483.

Seifritz E., Esposito F., Hennel F., Mustovic H., Neuhoff J.G., Bilecen D., Tedeschi G., Scheffler, K., Di Salle F., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297, 1706-1708.

Sekey A., 1963. Short-term auditory frequency discrimination. J. Acoust. Soc. Am. 35, 682-690.

Shah N.J., Jäncke L., Grosse-Ruyken M.L., Müller-Gärtner, H.W., 1999. Influence of acoustic masking noise in fMRI of the auditory cortex during phonetic discrimination. J. Magn. Reson. Imaging 9, 19-25.

Shah N.J., Steinhoff S., Mirzazade S., Zafiris, O., Grosse-Ruyken, M.L., Jäncke, L., Zilles, K., 2000. The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. Neuroimage 12, 100-108.

Shamma S.A., 1985. Speech processing in the auditory system. II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.* 78, 1622-1632.

Shamma S. & Klein D., 2000. The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. J. Acoust. Soc. Am. 107, 2631-2644.

Shamma S. A., Shen N. M., Gopalaswamy P., 1989. Stereausis: Binaural

processing without neural delays. J. Acoust. Soc. Am. 86, 989-1006.

Shannon R. V., 1976. Two-tone unmasking and suppression in a forward masking situation. J. Acoust. Soc. Am. 59, 1460-1470.

Siebert W.M., 1970. Frequency discrimination in auditory system: Place or periodicity mechanisms? Proc. IEEE 58, 723-730.

Singh N.C., Theunissen, F.E., 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. J. Acoust. Soc. Am. 114, 3394-3411.

Sivian L.J,. Dunn H.K., White S.D., 1959. Absolute amplitudes and spectra of certain musical instruments and orchestras. IRE Trans. Audio7, 47-75.

Smith P.H., Joris P.X., Yin T.C., 1993. Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive. J. Comp. Neurol. 331, 245-260.

Smith P.H., Joris P.X., Yin T.C., 1998. Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (MNTB) of the cat. J. Neurophysiol. 79, 3127-3142.

Soderquist D.R., Shilling R.D., 1990. Loudness and the binaural masking level difference. Bull. Psychon. Soc.28, 553-555.

Stecker G.C., Harring I.A., Middlebrooks J.C., 2005. Location coding by opponent neural populations in the auditory cortex. P.L.o.S. Biol. 3, e78.

Stern R.M., Shear G.D., 1996. Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. J. Acoust. Soc. Am. 100, 2278-2288.

Stern R.M., Trahiotis C., 1995. Models of binaural interaction, in *Hearing*, edited by

B.C.J. Moore. Academic, London, pp. 347 - 386.

Sullivan W.E., Konishi M., 1986. Neural map of interaural phase difference in the owl's brainstem. Proc. Natl. Acad. Sci. U.S.A. 83, 8400-8404.

Talairach J., Tournoux P., 1988. Co-planar stereotaxic atlas of the human brain. Thieme, Stuttgart.

Thompson S.K., von Kriegstein K., Deane-Pratt A., Marquardt T., Deichmann R., Griffiths T.D., McAlpine D, 2006. Reprensentation of interaural time delay in the human auditory midbrain. Nature Neurosci. 9, 1096 – 1098.

Tian B., Reser D., Durham A., Kustov A., Rauschecker J.P., 2001. Functional specialization in rhesus monkey auditory cortex. Science 292, 290-293.

Tollin D. J., 2003. The lateral superior olive: A functional role in sound source localization. Neuroscientist 9, 127-143.

Toole F.E., Sayers B. M., 1965a. Lateralization judgements and the nature of binaural acoustic images. J. Acoust Soc Am 37, 319-324.

Toole F. E., Sayers, B. M., 1965b. Inferences of neural activity associated with binaural acoustic images. J. Acoust. Soc. Am. 38, 769-779.

Tootell R.B., Hadjikhani N.K., Vanduffel, W., Liu, A.K., Mendola, J.D., Sereno, M.I., Dale, A.M., 1998. Functional analysis of primary visual cortex (V1) in humans. Proc. Nat. Acad. Sci. U.S.A. 95, 811-817.

Tootell R.B., Reppas J.B., Dale A.M., Look R.B., Sereno M.I., Malach R., Brady T.J., Rosen, B.R., 1995. Visual motion aftereffect in human cortical area MT revealedby functional magnetic resonance imaging. Nature 375, 139-141.

Townsend T.H., Goldstein D.P., 1972. Suprathreshold binaural unmasking. J. Acoust. Soc. Am. 51, 621-624.

Trahiotis C., Stern R.M., 1989. Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase. J. Acoust. Soc. Am. 86, 1285-1293.

Trussell L.O., 1999. Synaptic mechanisms for coding timing in auditory neurons. Annu. Rev. Physiol. 61, 477-496.

Turnbull W.W., 1944. Pitch discrimination as a function of tonal duration. J. Exp. Pscychol. 34, 302 - 316.

Ungan P., Sahinoğlu B., Utkaçal R., 1989. Human laterality reversal auditory evoked potentials: Stimulation by reversing the interaural delay of dichotically presented continuous click trains. Electorencephaolgr. Clin. Neurophysiol. 73, 306-321.

Ungan P., Yagacioglu S., Goksoy C., 2001. Differences between the N1 waves of the responses to interaural time and intensity disparities: Scalp topography and dipole sources. Clin. Neurophysiol. 112, 485-498.

van Bergeijk W.A., 1962. Variation on a theme of Békésy: A model of binaural interaction. J. Acoust. Soc. Am. 34, 1431-1437.

van de Par S., Kohlrausch, A., 1997. A new approach to comparing binaural masking level differences at low and high frequencies. J. Acoust. Soc. Am. 101, 1671-1680.

van de Par S., Kohlrausch A., Breebaart J., McKinney M., 2005. Discrimination of different temporal envelope structures of diotic and dichotic target signals within diotic wide-band noise in *Auditory signal processing : physiology,*

psychoacoustics, and models, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet. Springer, New York, pp. 398 - 404.

van der Heijden M., Joris P. X., 2006. Panoramic measurements of the apex of the cochlea. J. Neurosci. 26, 11462-11473.

Viemeister N.F., 1979. Temporal modulation transfer functions based upon modulation thresholds. J. Acoust. Soc. Am. 66, 1364-1380.

von Békésy G., 1930. Zur Theorie des Hörens. Über das Richtungshören bei einer Zeitdifferenz oder Lautstärkenungleichheit der beidersteitigen Schalleinwirkungen. Physik Z 31, 824-835.

Warren J.D., Griffiths T.D., 2003. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J. Neurosci. 23, 5799-5804.

Warren J.D., Zielinski B.A., Green G.G., Rauschecker J.P., Griffiths T.D., 2002. Perception of sound-source motion by the human brain. Neuron 34, 139-148.

Watkins K.E., Paus T., Lerch J.P., Zijdenbos A., Collins D.L., Neelin P., Taylor J., Worsley K.J., Evans A.C., 2001. Structural asymmetries in the human brain: A voxel-based statistical analysis of 142 MRI scans. Cereb. Cortex 11, 868-877.

Wiegrebe L., Patterson R. D., 1999. Quantifying the distortion products generated by amplitude-modulated noise. J. Acoust. Soc. Am. 106, 2709-2718.

Wightman F.L., Kistler D.J., 1989. Headphone simulation of free-field listening. I: Stimulus synthesis. J. Acoust. Soc. Am. 85, 858-867.

Wightman F.L ., Kistler D.J., 1992. The dominant role of low-frequency interaural time differences in sound localization. J. Acoust. Soc. Am. 91, 1648-1661.

Witton C., Green G.G., Rees A., Henning G.B., 2000. Monaural and binaural detection of sinusoidal phase modulation of a 500-Hz tone. J. Acoust. Soc. Am. 108, 1826-1833.

Yost W.A., Patterson R., Sheft S., 1996. A time domain description for the pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 99, 1066-1078.

Zatorre R.J., Penhume V.B., 2001. Spatial localization after excision of human auditory cortex. J. Neurosci 21, 6321-6328.

Zeng F.G., et al., 2004. On the dichotomy in auditory perception between temporal envelope and fine structure cues. J. Acoust. Soc. Am. 116, 1351-1354.

Zerlin S., 1969. Traveling-wave velocity in the human cochlea. J. Acoust. Soc. Am. 46, 1011-1015.

Zhou Y., Carney, L. H., Colburn, H. S., 2005. A model for interaural time difference sensitivity in the medial superior olive: Interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology. J. Neurosci. 25, 3046-3058.

Zwicker E., 1956. Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs. Acustica 6, 365-381.

Zwicker E., 1970. Masking and psychological excitation as consequences of the ear's frequency analysis, in *Frequency analysis and periodicity detection in hearing* edited by R. Plomp, and G.F. Smoorenburg. A. W. Sijthoff, Leiden, pp. 376-396.