The role of temporal and spectral cues in the temporal integration of pitch and in pitch-based segregation of sound sources

Nicholas R. Clark, BSc.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy, July 2011

ABSTRACT

The auditory nerve conveys spectral information, reflecting the location of maximum vibration along the frequency-tuned basilar membrane, and also information reflecting the timing of peaks in the vibrations at each location. Debate continues as to whether pitch is extracted based on the available temporal or spectral representations of tonal stimuli, or both. The aim of the current work was to determine the roles of temporal and spectral harmonicity cues for pitch, under important conditions for understanding speech in multi-talker environments. Two such conditions are the temporal integration of pitch and pitch-based segregation of sound sources.

Pitch information in running speech changes over time. Therefore, pitch-extraction mechanisms must be able to follow these changes to enhance intelligibility, particularly when listening in modulated backgrounds such as competing speech. However, the temporal resolution of pitch has received little attention. In the first three chapters, the roles of temporal and spectral cues on the temporal resolution of pitch extraction were determined by measuring pitch-domain temporal modulation transfer functions and gap-detection thresholds. Temporal resolution was shown to be unaffected by the availability of spectral cues, and similarly unaffected by the overall pitch strength of the stimulus. However, the system was much more sluggish in response to changes in pitch information in stimuli presented in high-frequency regions compared to low-frequency regions. This processing strategy may reflect the progressive loss of accurate temporal information towards higher frequencies imposed by transduction processes in the auditory periphery.

To understand speech in noise, the ability of the auditory system to integrate pitch information over long periods is equally important as its ability to detect rapid changes in pitch. In Chapter 4, discrimination thresholds for pitch value and pitch strength were measured in the presence and absence of spectral cues as a function of stimulus duration.

The assumption was that discrimination thresholds would reach asymptote at the stimulus duration corresponding to the length of the pitch integration window. However, the pitch-strength discrimination data revealed integration was only limited by the stimulus duration, suggesting that this task may reflect the rate of decrease in the variance of internal pitch-value and pitch-strength estimates with increasing stimulus duration, but not the total integration capacity of the system.

In multi-talker environments, listeners have to process multiple simultaneous tonal sound sources. The fifth study showed that temporal interactions between simultaneous tonal stimuli could aid detection in the absence of spectral cues. In contrast, harmonic resolvability is thought to be a prerequisite for pitch-based simultaneous grouping. However, data from a second experiment showed that listeners were able to perceptually segregate tonal sounds in the absence of spectral cues.

ACKNOWLEDGEMENTS

Firstly I would like to give thanks my supervisor Dr. Katrin Krumbholz for all of the time invested in me and her willingness to answer my questions at a moment's notice.

Special thanks to Charlotte Breakey who helped to collect the data presented in Chapters 1 – 4. I would also like to thank Tom Walters for helping me to get to grips with his AIM-C software used to model data in chapters 3 and 5. Thanks also to my second supervisor, Dr. John Crowe for reading reports and providing helpful feedback.

The memories I will take away from Nottingham will be of the times spent with the friends I have made here. My office mates past and present, David Magezi, Paul Briley, Barrie Edmonds, Gemma Hutchinson and Cris Lanting have always been kind and helpful. Many thanks to Sam Irving, Jazz and Calum Grimsley, Becky and Kyle Nakamoto, Simon Jones, Mike Pilling, Chris Scholes, Darren Edwards, Joe Sollini, Daphne Garcia, Jess Monaghan, Damon McCartney, Stefan Kerber, Jonathan and Sophie Laudanski, Kerri and Matt Young, Ian Wiggins, Heather Gilbert and Ben Irving for being great fun to be around both in and out of work.

My parents, in-laws, family, and friends back home have been great through this process. I'm, very lucky to know that I can always count on you and have your support in whatever I decide to do.

Finally, I would like to acknowledge my wife Andrea. Your strength and patience over the past few years, through good times and bad, has been an inspiration. I cannot begin describe my gratitude in words. With you, I have everything.

SUMMARY OF MAIN FINDINGS

The size of the temporal integration window used by the human auditory system for the neural extraction of pitch information is quantified in Chapters 1-3. This was achieved by measuring the acuity with which listeners were able to detect changes in serial correlation of a monaural stimulus (perceived as changes in pitch strength) over time under different listening conditions. Data from listening tests and simulations from computational models of auditory function suggested that:

1. The harmonic resolvability by the cochlear filters of individual frequency components of the stimuli has no significant effect on the duration of the integration window.
2. The duration of the integration window is inversely proportional to the repetition rate (perceived as pitch value) of the stimulus.
3. The auditory system is equally sensitive to changes in temporal regularity in stimuli presented in different spectral regions. However, the temporal acuity of the pitch-extraction mechanism is poorer in high-frequency regions compared to in low-frequency regions.
4. The size of the integration window is not dependent on the average pitch strength of the stimulus, but the auditory system is less sensitive to pitch-strength modulations in stimuli with weak pitch strength compared to stimuli with more salient pitch.
5. Established autocorrelation-based models of pitch perception can be modified to quantitatively account for all of the experimental observations.

Chapter 4 described a study that measured the total duration over which the brain is able to accumulate pitch information. This was achieved using paradigms that measured either pitch-value or pitch-strength discrimination thresholds as a function of the stimulus duration.

1. The integration times inferred from the data were far longer for the pitch-strength discrimination task than for the pitch-value discrimination task.
2. The result was qualitatively accounted for by using a model based on signal detection theory, comparing the pitch-value or pitch-strength resolution of the auditory system with the variance in the physical stimulus property responsible for each percept.

In the final Chapter, a pair of simultaneous tonal stimuli was used to investigate the role of pitch cues in detection and sound-source segregation. It was shown that:

1. Detection of a tonal signal in the presence of a tonal masker was facilitated by a reduction in the correlation (heard as a reduction in pitch strength) of the composite stimulus when the signal was present, relative to when the masker was presented alone. While the masking patterns for resolved and unresolved stimuli were different, an autocorrelation-based model of pitch was able to account for the experimental observations with very high accuracy.
2. There was a large effect of harmonic resolvability observed in an experiment where the listeners had to use pitch cues in order to perceptually segregate competing sounds to perform the task. However, the data suggested that harmonic resolvability is not a prerequisite for simultaneous sound-source segregation based on pitch cues. Listeners were able to separate spectrally unresolved auditory objects in the acoustic mixture, given a large enough pitch-value difference between the components.

LIST OF ABBREVIATIONS

2I2AFC	two-interval two-alternate forced-choice
3I3AFC	three-interval three-alternate forced-choice
A	amplitude
ACF	autocorrelation function
AIM	auditory image model
AM	amplitude modulation
BMLD	binaural masking level difference
C	criterion
d	delay
D	decision measure
dB	decibel
DC	direct current
E	expansive process
ERB	equivalent rectangular bandwidth (Glasberg and Moore, 1990)
F0	fundamental frequency
FFT	fast Fourier transform
f_m	modulation frequency
f_c	centre frequency
g	gain
GTFB	gammatone filter bank
$H1_{NAP}$	average first peak height in the autocorrelogram of the NAP
$h1_{NAP}$	instantaneous first peak height in the autocorrelogram of the NAP
$H1_S$	average first peak height in the autocorrelogram of the signal
$h1_S$	instantaneous temporal regularity

HCT	harmonic complex tone
Hz	hertz
i	peak index
I/O	input/output
IRN	iterated rippled noise
IRNO	iterated rippled noise add-original configuration
ITD	interaural time difference
j	imaginary unit
$\underline{k}$	expansive constant
kHz	kilohertz
$L1_{NAP}$	average first peak lag in the autocorrelogram of the NAP
$L1_S$	average first peak lag in the autocorrelogram of the signal
m	modulation index
MR	mix ratio
MRC	Medical Research Council
ms	milliseconds
n	number of iterations in IRN circuit
NAP	neural activity pattern
PZFC	pole zero filter cascade
Q	quality factor
R^2	correlation coefficient
R	normalized running integration process
R_{gap}	gap rate (inverse of T_{gap})
R_{mod}	modulation rate (inverse of T_{mod})
RIN	regular interval noise

RPN	repeated period noise
RMS	root mean square
RN	rippled noise
SPL	sound pressure level
SMR	signal to masker ratio
SNR	signal to noise ratio
t	time
TDT	Tucker-Davies Technologies
TI	time interval
T_{gap}	gap duration
T_{mod}	modulation period
TMTF	temporal modulation transfer function
ΔI	signal intensity difference between observation intervals
μs	microseconds
η	pitch integration time constant scalar
ω	angular frequency
ϕ	start phase

TABLE OF CONTENTS

GENERAL INTRODUCTION

The most important role of the human auditory system is as a receiver for speech communications. The pulsation of the vocal chords at regular intervals during the production of voiced speech gives the speech a harmonic structure. This harmonic structure is extracted by the central auditory system, giving rise to a pitch associated with the speech. Pitch is generally thought of as the perceptual attribute associated with musical melodies. However, in speech, pitch conveys prosodic information, such as whether an utterance is a question or a statement. In tonal languages, such as Mandarin, pitch even contains phonological information. Pitch also conveys information about speaker identity.

In situations with multiple speakers, it is unlikely that each speaker will concurrently produce speech with identical glottal pulse rates. Therefore, the pitch of each speaker's voice can be extracted by the auditory system and used as a cue for grouping information from the acoustical environment and assigning it to individual sources. Scheffers (1983) was the first to use a simultaneous vowel paradigm to quantify segregation performance based on pitch cues. Listeners were presented with two simultaneous vowel sounds and asked to identify each. Performance increased markedly when a small rate difference was introduced between the vowels. This effect has been reliably replicated in numerous studies (Zwicker, 1984, Assmann and Summerfield, 1989, Assmann and Summerfield, 1990, Culling and Darwin, 1993). Incidentally, hearing-impaired listeners have considerable difficulty when listening in backgrounds of competing speech. Hearing-impaired listeners perform somewhat more poorly than normal-hearing listeners when listening in the presence of steady background sounds, but perform considerably more poorly when listening in the presence of modulated background sound (Duquesnoy, 1983). Normal-hearing listeners are able to exploit the signal information that is revealed in the low-amplitude segments of background sounds –

a strategy known as dip listening. Conversely, hearing-impaired listeners have little or no ability to utilize information within the dips, even when sounds within the dips are amplified to be above absolute threshold (Moore et al., 1999). Therefore, not only is it important to understand how pitch is extracted from a simple tonal sound, but it is especially important to understand how it is extracted in complex and highly dynamic stimuli exhibiting the features of speech.

Before one can hypothesize about how the brain extracts pitch information from an acoustic stimulus, one must first have knowledge of the information conveyed to the central auditory system via the auditory nerve. Within the cochlea, the basilar membrane vibrates sympathetically with the temporal waveform of the stimulus. The mechanical properties of the membrane are such that regions near the base respond maximally to high-frequency spectral components, while regions near the apex respond maximally to low-frequency components. Information along the length of the membrane is transferred by individual nerve fibres, giving a place, or spectral coding, of the stimulating sound. If the sound frequency is not too high, the action potentials are time-locked to the individual basilar membrane deflections within each spectral channel. The resultant timing information is referred to as temporal fine structure (TFS). In mathematics, both time- and frequency-domain representations of a signal are identical in terms of the information that they contain and are related via the Fourier transform. Within the auditory system, the information conveyed by spectral and temporal representations of the stimulating sound are not equivalent because the auditory periphery imposes unique limitations upon each representation.

Each place along the basilar membrane behaves like a band-pass filter, attenuating frequency components away from its best frequency. Each filter has a relatively constant quality factor (Q); therefore, the bandwidth of each filter increases with its best frequency.

Harmonic signals contain energy at integer multiples of their fundamental frequency. The spectral resolution of the auditory filters in hearing-impaired people is reduced relative to normal-hearing people (Glasberg and Moore, 1986). Therefore, the accuracy of the spectral representation of the stimulus is even further reduced. A harmonic signal with spectral energy distributed across the range of human hearing may be accurately represented by auditory spectral coding at relatively low frequencies. Here, individual partials maximally activate distinct spatial regions and are said to be resolved. However, at higher frequencies, many harmonic components are likely to fall within the passband of individual filters. This gives a flat internal spectral representation, and the individual partials are said to be unresolved. Therefore, the spectral resolvability of a harmonic stimulus can be controlled by independent adjustment of its fundamental frequency and the spectral band in which it is presented, thus giving the experimenter control over the spectral information available to the central auditory system.

The accuracy of the temporal representation of the stimulus is primarily limited by the mechanical-to-neural transduction process. There is a phase-locking limit to which this process is readily able to transmit the timing of peaks in the fine structure to the central auditory system. This is due to both the inner hair cell (IHC) membrane time constant (Palmer and Russell, 1986) and jitter in transmission at the synapses between the IHC and primary neurons (Anderson et al., 1971). In humans, the breakdown in phase locking is thought to occur between 0.8 and 1.2 kHz, above which the reliability of TFS information is degraded. However, information about the slowly fluctuating amplitude envelope of the signal within the filter can still be transmitted. Not only do hearing-impaired listeners have a reduced ability to resolve individual frequency components of complex sounds, but it has also more recently been suggested that hearing-impaired listeners have a reduced ability to process TFS information (Lorenzi et al., 2006).

The fact that both temporal and spectral representations of tonal stimuli are transmitted to the brain has led to the development of corresponding temporal and spectral models of pitch extraction. Pitch is evoked by stimuli that are periodic, and pitch value depends primarily on the period of the stimuli. Both spectral and temporal pitch-extraction models share the common goal of extracting the periodicity from the stimulus, even when the fundamental partial is missing. Spectral pitch-extraction models are generally based on pattern matching. This involves analysis of the distribution of peaks in the internal spectrum of the stimulus. The brain is exceptionally good at recognising patterns from sensory inputs and also at perceptually reconstructing missing parts. Pattern-matching models assume that this is how pitch is perceived when the fundamental partial of the harmonic series is missing. The most well-known of these are the closely related models of Goldstein (1973), Wightman (1973) and Terhardt (1974). Temporal pitch-extraction models are generally based on an autocorrelation-type process that analyses the temporal regularity of the firing patterns present in auditory nerve fibres. This type of model was originally proposed by Licklider (1951) and was later reformulated and implemented computationally (Meddis and Hewitt, 1991a, Meddis and Hewitt, 1991b). Other well-known models based on the same principles include the Equalisation Cancelation model (de Cheveigné, 1998) and the Strobed Temporal Integration model (Patterson, 1994, Patterson and Irino, 1998).

Listeners can perceive pitch in both resolved and unresolved harmonic stimuli. While spectral pattern-recognition models can only extract pitch information from stimuli containing resolved harmonics, autocorrelation-type models have the distinct advantage of being able to extract pitch information from unresolved as well as resolved stimuli. However, behavioural studies have shown marked differences between performance in pitch discrimination (Houtsma and Smurzynski, 1990, Carlyon and Shackleton, 1994,

Carlyon, 1996b, Carlyon, 1998) and segregation tasks (for review, see Darwin and Carlyon, 1995) comparing resolved and unresolved harmonic stimuli, thus suggesting coexistence of both spectral and temporal pitch-extraction mechanisms.

The ability of the auditory system to detect relatively rapid changes in sounds over time is essential for understanding dynamic sounds such as speech. However, there is much variation in the temporal resolution of the auditory system in response to changes in different sound attributes. In psychoacoustical literature, it is customary to present the contrasting metaphorical images of the monaural hare and the binaural slug. This refers to the contrast between the excellent temporal resolution of the monaural auditory system in response to changes in intensity, compared to the inability of the system to detect fast changes in binaural parameters. The peripheral processing of both monaural and binaural signals is identical; therefore, the sluggishness must arise from differences in the central integration processes involved in extracting information. However, this comparison is based on just one attribute of a monaural signal: its intensity. Conversely, the temporal resolution of the monaural auditory system in response to pitch information has received little attention (Wiegrebe, 2001).

The general aim of this work was to investigate how pitch is extracted from complex and highly dynamic stimuli exhibiting the features of speech. The temporal dynamics of pitch perception were measured, as the ability of the listener to hear changes in pitch over time is crucial for following the running speech of individual talkers. The use of pitch cues for detection and segregation of simultaneous sound sources was also measured in order to understand how multiple pitches are extracted simultaneously.

In **Chapter 1** of this thesis, the temporal resolution of the neural pitch-extraction mechanism was measured using a novel stimulus that allowed for experimental analogues of classical intensity envelope resolution paradigms, such as gap detection (Fitzgibbons

and Wightman, 1982, Plomp, 1964) and modulation detection (Viemeister, 1979) to be conducted in the pitch domain. Resolution was measured in the presence and absence of spectral cues by using both resolved and unresolved stimuli. The results of this study suggest that harmonic resolvability has no effect on the temporal resolution of the pitch-extraction mechanism.

In the first chapter, stimuli were presented in a listening region where some of the spectral energy was within the putative phase-locking range of the mechanical-to-neural transduction process. Therefore, TFS information was readily available for both resolved and unresolved stimuli. In **Chapter 2**, the effect of harmonic resolvability on the temporal resolution of the pitch-extraction mechanism was measured using stimuli that were presented in a high-frequency band. While the exact limit of human phase-locking is unknown, the fidelity of the TFS within the high-frequency band would have been expected to be severely degraded relative to that in the low-frequency band. Therefore, a temporal pitch extraction mechanism would have less information to work from, and so an effect of harmonic resolvability may have been more likely to manifest itself in this band. However, the results of this study showed that while the temporal resolution of the pitch extraction mechanism was more sluggish in the high-frequency band relative to the low-frequency band, there was still no effect of harmonic resolvability.

The frequency region in which a tonal stimulus is presented is known to have an effect on its subjective pitch strength. The data from Chapter 2 suggested that the pitch extraction mechanism is more sluggish in a higher-frequency region. Therefore, the temporal resolution of the auditory system may be dependent on the pitch strength associated with the stimulus. This was tested in **Chapter 3** by varying the pitch strength of the stimulus directly, rather than changing the frequency region in which it was presented. The results suggested that the temporal resolution of the auditory system is invariant with

pitch strength. However, results suggested that listeners were more sensitive to changes in pitch strength in stimuli with a higher overall subjective pitch strength. The second part of this chapter considered the implications of cochlear compression on how sensitivity should be modelled in a neural model of pitch strength.

As with other senses, in audition, detection and discrimination performance generally improve with increasing stimulus duration. Therefore, to understand speech in noise, the ability of the auditory system to integrate pitch information over long periods is equally important as its ability to detect rapid changes in pitch cues. Performance in pitch-value discrimination tasks generally improves with increases in duration up to a point, after which thresholds no longer improve with further increases in duration. This has generally been taken to reflect the integration capacity of the system, and studies have shown that duration over which the system is able to integrate pitch information is dependent on the availability of spectral cues (Plack and Carlyon, 1995, White and Plack, 1998). However, data from Chapters 1 – 3 showed that the temporal resolution of the pitch extraction mechanism was not dependent on the resolvability of the stimuli. Furthermore, the duration of the pitch-integration windows measured in a more recent study (White and Plack, 2003) were shorter than those required to explain the sluggishness of the pitch-extraction mechanism observed in Chapter 2. This paradoxical comparison was investigated in two separate experiments in **Chapter 4** by measuring the effects of stimulus duration on the discriminability of pitch cues. In the first of these experiments, pitch-value discrimination thresholds were measured, and in the second of these experiments, pitch-strength discrimination thresholds were measured. If the effects of stimulus duration can be used to infer the integration capacity of the system, then one would have expected to see similar effects of stimulus duration in both experiments. However, results from the pitch-strength discrimination experiment indicated that

integration was being performed over much longer durations than indicated by the results from the pitch-value discrimination experiment. A model was presented to account for these different results in terms of the variance of the simulated internal representations of pitch-strength and pitch-value.

The data from Chapters 1 – 4 showed no effects of harmonic resolvability on integration times. However, spectral differences between signal and masking sounds are known to aid the detectability of a signal. Furthermore, pitch differences between simultaneous sounds are known to play an important part in grouping (Darwin, 1981), allowing the listener to perceptually segregate the two sources. However, this is only thought to be the case when spectral cues are available (Micheyl et al., 2006).

In **Chapter 5**, the role of pitch cues in aiding detection and segregation of simultaneous tonal sound sources was investigated. This is especially important for understanding how pitch cues aid intelligibility in multi-talker environments. Recently, it has been shown that pitch cues can be used to aid detection of a tonal signal in the presence of a tonal masker based on the temporal interactions between the competing sounds (Krumbholz et al., 2003a). In the first experiment presented in the chapter, the results of Krumbholz and colleagues (2003a) were extended by including harmonic resolvability as an experimental parameter. While there was an effect of harmonic resolvability, a temporal model of pitch extraction was nevertheless able to account for the almost all of the observed masking release. In contrast with the first half of the chapter, the second half measured the ability of the auditory system to segregate simultaneous sound sources based on pitch cues using a novel paradigm. Contrary to common opinion, observations from the second half of the study suggested that harmonic resolvability is not necessarily a prerequisite for pitch-based segregation.

Chapter 1

The temporal resolution of pitch perception I: The Monaural Slug

I. INTRODUCTION

Hearing impairment is commonly associated with a reduction in sensitivity to sounds, which leads to an increase in detection thresholds. However, the biggest problem for hearing-impaired listeners is understanding speech in noisy environments. Hearing-impaired listeners perform somewhat more poorly than normal-hearing listeners in the presence of steady background sounds, but are known to have particular difficulties in understanding speech in modulated backgrounds, compared to normal-hearing listeners (Duquesnoy, 1983), and this may be the result of a deficit in temporal processing ability. Pitch information is one of the most important cues for hearing out individual talkers in such environments (Zwicker, 1984, Assmann and Summerfield, 1989, Assmann and Summerfield, 1990, Culling and Darwin, 1993); however, the temporal dynamics of pitch perception are relatively poorly understood in comparison with the temporal dynamics of other sound attributes such as intensity and binaural cues.

Early studies on temporal resolution in the auditory system generally refer to its ability to track changes in the intensity envelope of a sound. A common experimental paradigm used for this is the gap-detection task (Fitzgibbons and Wightman, 1982, Plomp, 1964). For this, listeners are typically asked to detect a brief decrement in the intensity of a sound. Detectability of the gap generally increases with gap duration. Another common paradigm is amplitude modulation detection (Viemeister, 1979), where detectability of the modulation decreases with rate. The success of modulation- and gap-detection paradigms in quantifying envelope resolution has led to the development of analogues of these paradigms for use in quantification of binaural temporal resolution (Grantham, 1982, Akeroyd and Summerfield, 1999). Limitations in temporal resolution are usually attributed to a central integration process that integrates information over a temporal window, thus reducing the dynamic range of fluctuations in the internal representation of the sound. The

limitations imposed by these central integration processes are highly dependent on the information being integrated; for example, resolution for binaural processing has been measured to be around two orders of magnitude slower than intensity resolution (Grantham, 1984, Grantham, 1982, Grantham and Wightman, 1978).

Temporal regularity within an auditory stimulus gives rise to the perception of pitch. The temporal dynamics of pitch have mainly been investigated in tasks measuring the duration over which the auditory system is able to integrate information in order to improve performance in pitch-discrimination tasks (Krumbholz et al., 2003b, Plack and Carlyon, 1995, White and Plack, 1998, White and Plack, 2003). In general, the results from these studies suggest that the duration of the integration window for pitch depends on the harmonic resolvability of the stimulus, and this has been taken to suggest that the pitch of resolved and unresolved harmonic complex tones are extracted by different mechanisms. However, the temporal resolution of pitch extraction has received little attention.

Temporal models of pitch extraction assume that an autocorrelation-like process is responsible for analysis of the periodicity within the firing patterns conveyed by auditory nerve fibres. To be able to detect changes in pitch information over time, this process must be calculated within a finite-duration window that shifts along the time axis. Licklider (1951) was the first to suggest that an autocorrelation process may be responsible for pitch extraction in humans. He suggested that pitch information may be integrated over an exponentially decaying window with a time constant of 2.5 ms. The integration window acts like a moving- average filter, and so the longer the window, the more it attenuates rapid fluctuations in pitch information. Until relatively recently, the time constant used in computational realisations (Meddis and Hewitt, 1991a, Meddis and Hewitt, 1991b, Meddis and Omard, 1997) of Licklider's model had been treated as a free parameter.

Only two studies have measured the temporal resolution of pitch extraction (Balaguer-Ballester et al., 2009, Wiegrebe, 2001). Both studies used a class of stimuli called regular-interval noises (RIN) that are derived from random noise but contain some temporal regularity within the waveform; therefore, they give rise to a 2-component perception consisting of a buzzy pitch and a noise. Wiegrebe (2001) used a subcategory of RIN known as repeated-period noise (RPN), which was generated by concatenating identical noise samples of duration d. By periodically introducing uncorrelated noise samples into the sequence, the temporal regularity within the stimulus was switched on and off. Therefore, when the modulations were slow enough, listeners heard a sound with a pitch strength that switched between that of a Gaussian noise and that of a random-phase harmonic complex with fundamental frequency of 1/d. Wiegrebe (2001) was unable to account for his results using a model with a single integration time constant and proposed that the size of the temporal window depends on the pitch itself.

Balaguer et al. (2009) also conducted an experiment to assess the temporal resolution of pitch perception. For this, a different type of RIN stimulus was used, called ripple noise (RN). This was generated by delaying a Gaussian noise sample by a delay, d, and adding the delayed copy back to the original. Like RPN, RN has temporal regularity that gives rise to a pitch percept. The pitch value of RN corresponds to the reciprocal of the delay. This temporal regularity can be switched on and off over time by replacing portions of the delayed noise copy with an independent Gaussian noise. As in Wiegrebe's (2001) study, the RN stimulus was used in an experiment where the detectability of square-wave modulations in pitch strength was measured, and also in an experiment where the detectability of a single gap in pitch strength was measured. Unlike the stimuli used by Wiegrebe (2001), where pitch-strength modulation rates were limited to integer multiples of d, the modulation rates used in Balaguer (2009) were adjustable as a continuum.

Thresholds were measured for the shortest detectable gap and the fastest detectable modulation rate in RNs, where d was equal to 1, 2, 4, 8, and 16 ms. The thresholds measured in this study are shown in Fig. 1. Both gap- and modulation-detection thresholds increase with delay, suggesting that the temporal resolution of the pitch-extraction mechanism is higher for higher-pitched (shorter d) stimuli.

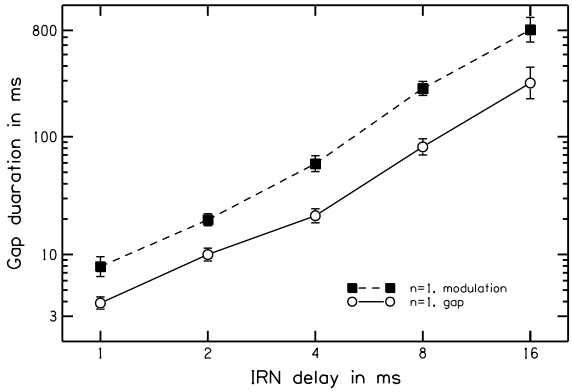

FIG. 1. Data re-plotted from Balaguer et al. (2009), where each threshold shown is averaged across five listeners and the parameter is the detection task. The ordinate shows the delay, d, used in the RN circuit, and the abscissa shows the gap duration, where gap duration is the length of the uncorrelated noise sample inserted into the delayed path of the RN. Error bars represent the inter-listener standard error.

Results from Balaguer et al. (2009) were single-value measures of temporal resolution, in that thresholds were measured for the shortest detectable gap in temporal regularity without adjusting the depth of the gap. Buunen and van Valkenburg (1979) showed how the shortest detectable gap in stimulus intensity was dependent on the depth of the gap. In single-value measures of resolution, the degree of smearing from the integration process cannot be disentangled from the sensitivity of the auditory system to the modulations. Therefore, single-value measurements of temporal acuity do not provide

enough information to quantitatively determine the time constants of the integration process directly from the data.

To overcome the limitations imposed by existing stimuli, the current experiments used a novel stimulus where the instantaneous temporal regularity could be adjusted to any desired value as a function of time. The new stimulus permitted measurement of the temporal resolution of pitch in a gap-detection task, where the gap depth could be adjusted independently of its duration. Thresholds were measured for the smallest detectable reductions in pitch strength for finite-duration gaps placed at the temporal centre of the stimulus. The novel stimulus also permitted measurement of pitch-domain temporal modulation transfer functions (TMTFs), which are a measure of how a system responds to sinusoidal modulations in pitch strength at different modulation rates. This approach was particularly attractive, because if the system is linear in response to fluctuations in serial correlation, then the TMTF measurements are able to predict the output of the system in response to an arbitrary input. Therefore, no assumptions about the underlying pitch-extraction processes need to be made in order to derive the time constants of pitch extraction. Comparison of results from gap- and modulation-detection tasks allowed for determination of whether the time constants of pitch perception are task-dependent. The harmonic resolvability of the stimuli was also included as an experimental parameter so that results could be directly compared with those from pitch-integration studies.

II. THE NOVEL STIMULUS

The temporal regularity and thus the perceived pitch strength associated with a RN stimulus can be increased by iterating the delay-and-add process n times to produce iterated ripple noise (IRN). One way to achieve this is by summing the signal present in the delay line with the original signal after each iteration: IRN add-original (IRNO) (Yost,

1996). The autocorrelogram of an IRN consists of a series of peaks at integer multiples of d. Yost (1996) showed that by subjecting the stimulus to an autocorrelation process integrated across the stimulus duration, the pitch strength associated with an IRN is monotonically related to the height of the peak occurring at a lag equal to 1d in the autocorrelogram. The height of the peak in the autocorrelogram of the stimulus occurring at 1d ($H1_S$) is dependent on n and can be determined analytically, as shown in Eqn. 1.

$$H1_S = \frac{n}{n+1} \qquad \text{(EQN. 1.)}$$

The subscripted S in $H1_S$ is to denote that the autocorrelation is performed directly on the stimulus, as opposed to a simulated pattern of auditory nerve firing, which is discussed later in this chapter. In general, IRNs give rise to temporally invariant pitch. However, it is possible to modify an IRN circuit to facilitate modulation of the delay over time (Denham, 2005), resulting in a temporally dynamic percept of pitch value. Inspired by this modification, a novel IRN circuit was created (Fig. 2), facilitating modulation of the temporal regularity within the IRN over time. In the modified circuit, a new noise from an uncorrelated source was introduced at each iteration and then mixed with the signal present in the delay line according to the ratio determined by g(t). This gave rise to a pitch percept that could be varied anywhere between that of a noise and that of an IRN as a function of time. In the modified circuit used here, the instantaneous temporal regularity ($h1_S$) at a given point in time is defined as a function of n, and the dynamic gain parameter, g(t), as shown in Eqn. 2.

$$h1_S = \frac{n}{n+1} g(t) \qquad \text{(EQN. 2.)}$$

Here, the h in $h1_S$ is printed lowercase to differentiate between the instantaneous first peak height and the first peak height calculated by subjecting the signal to an autocorrelation integrated across the stimulus duration, $H1_S$.

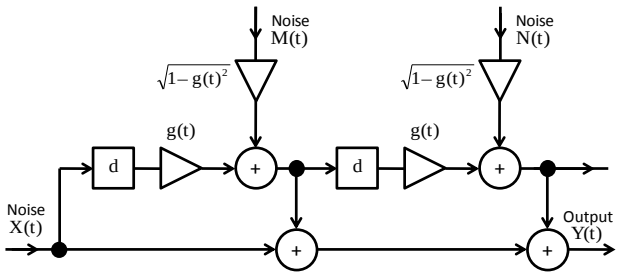

FIG. 2. Signal-flow diagram showing two iterations of the modified IRNO algorithm that allows temporal regularity to be varied over time by modulation of g(t). Within each iteration block, an uncorrelated noise signal is introduced with power reciprocal to that of the signal present in the delay line. This ensures that the total power output of the stimulus remains constant over time, irrespective of instantaneous changes in temporal regularity.

III. METHODS

A. Stimuli

In the current experiments, a total of 16 iterations of the dynamic IRN circuit were used to give a large potential dynamic range of gap depth and modulation index. To investigate interactions between integration time and stimulus repetition rate, thresholds were measured for 4 different repetition rates spanning 2 octaves around 75 Hz. IRN repetition rates were 1 octave above (150 Hz) and below (37.50 Hz) and 0.5 octaves above (106.07 Hz) and below (53.03 Hz) a central value of 75 Hz. IRNs have harmonic comb spectra, the harmonic resolvability of which was an experimental parameter. Harmonic resolvability is defined according to the rule of Shackleton and Carlyon (1994): when fewer than 2 harmonics are present in the 10-dB bandwidth of the auditory filter, the

excitation pattern is taken to be resolved; when 2 to 3.25 harmonics are present in the 10-dB bandwidth, the excitation pattern is in a state of partial resolvability; and when more than 3.25 harmonics fall within the 10-dB bandwidth, the excitation pattern is unresolved. The number of harmonics in the 10-dB bandwidth of an auditory filter was estimated as the repetition rate of the IRN divided by 1.8 times the equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990). To measure the effects of the harmonic resolvability, stimuli were filtered into a 2.2-kHz bandwidth with a centre frequency of 1.88 kHz. The lower cutoff frequency of the spectral band (0.78 kHz) was set to coincide with the mean value of harmonics per 10-dB bandwidth to achieve partial resolvability (2.625) at a repetition rate of 75 Hz. According to this rule, the 2 lower rates (37.50 Hz, 53.03 Hz) were entirely unresolved, whereas the 2 higher rates (106.07 Hz, 150.00 Hz) contained at least some resolved components within the lower part of the band.

Stimuli were presented at a level of 65-dB sound pressure level (SPL) and were gated on and off with 5-ms cosine-squared ramps to prevent audible clicks at the onset and offset of stimulus intervals. Stimuli were presented in a continuous noise to mask audible distortion products below the stimulus passband. This noise was lowpass filtered at 0.5 octaves below the lower cutoff frequency of the stimulus passband using an 8th order Butterworth filter. Prior to lowpass filtering, the noise was filtered in the spectral domain so as to produce a roughly constant excitation level of 50-dB SPL per equivalent rectangular bandwidth.

Stimuli were generated digitally with a sampling rate of 25 kHz and digital-to-analogue converted with a 24-bit resolution using MATLAB (The Mathworks, Natick, MA, USA) and the real-time processor (TDT RP 2.1) of TDT System 3 (Tucker-Davies Technology, Alachua, FL, USA). They were 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 (IAC, Winchester, UK).

B. Procedure

Gap-detection thresholds were measured for gap durations (T_{gap}) equal to multiples of 1, 2, 4, 8, 16, and 32 times each IRN delay, d. Informal listening showed that the modulation-detection experiment was more difficult and required slower modulation rates to achieve a good dynamic range of thresholds. Therefore, modulation-detection thresholds were measured for modulation periods (T_{mod}) equal to multiples of 3, 6, 12, 24, 48, and 96 times each IRN rate. The longest T_{mod} used for the 37.50-Hz IRN was limited to 48 times the respective d, as T_{mod}=96d would require a stimulus duration in excess of 2.5 seconds to capture a single modulation cycle. The stimulus duration was set to a factor of $\sqrt{2}$ longer than the longest respective T_{gap} or T_{mod}; therefore, durations of 1.2068 seconds were used in the gap-detection experiment and durations of 1.8102 seconds were used in the modulation-detection experiment.

Each trial consisted of three observation intervals, which were separated by 500-ms gaps. Two intervals contained unmodulated stimuli, while the remaining interval contained the target stimulus with the modulated $h1_S$. Intervals were presented in a random order within each trial. In the target intervals of the gap experiment, the gap was positioned symmetrically around the temporal centre of the stimulus. In the target intervals of the modulation experiment, the modulation was presented continuously throughout the stimulus with random start phase.

Gap depths were manipulated by setting g(t) equal to 1 for the duration of the stimulus, apart from in the region of the gap, where it was set to 1–m, where m is the gap depth. Therefore, a gap depth of 0 dB gave an $h1_S$ of 0 in the gap region (full depth). A

gap depth of -6 dB corresponded to a gap where the $h1_S$ in the gap region was halfway between 0 and the maximum. In the modulation experiment, modulation was introduced by setting g(t) according to Eqn. 3, where m is the modulation index, f_m is the modulation rate, and ϕ is the random starting phase. The modulation index used was equivalent to that used in amplitude modulation, but was normalized to values between 0 and 1, because g needs to be in the range of 0 to 1.

$$g(t) = \frac{1 + m\cos(2\pi f_m t + \phi)}{2} \qquad \text{(EQN. 3.)}$$

In a standard amplitude-modulation (AM) detection task, the listener must discriminate between a modulated signal and an unmodulated signal. In an AM detection task, the root-mean-square (RMS) levels of both intervals are equal to prevent overall loudness cues. Due to the nonlinear relationship between $h1_S$ and the perceived pitch strength (Yost, 1996), some different precautions are required when modulating pitch strength. When the modulation rate of $h1_S$ is increased above the modulation-detection threshold, it is perceived as having a static pitch salience. Pilot testing showed that perceived pitch strength of an IRN with $h1_S$ modulated at a rate above detection threshold was greater than that of an unmodulated IRN, even though both stimuli had an equal $H1_S$. This pitch-strength asymmetry was also reported by Wiegrebe et al. (1998). To ensure that listeners based their decisions on modulation detection and not salience discrimination, the overall salience cues had to be neutralized. One possible solution would have been to rove the pitch strength of the IRNs in non-target intervals. However, a more elegant solution was reached by matching the pitch-strength of modulated and unmodulated intervals. Wiegrebe et al. (1998) showed that the pitch-strength differences between modulated and unmodulated stimuli could be accounted for by an expansive process, E (similar to that shown by Eqn. 4).

$$E(h1_S) = \frac{10^{k*h1_S} - 1}{10^k - 1} \qquad \text{(EQN. 4.)}$$

The expanded $E(h1_S)$ is proportional to the pitch strength associated with the stimulus, where k is a constant that controls the amount of expansion. Pitch-strength cues between intervals were neutralised in the current study by adjusting the $E(H1_S)$ of the unmodulated intervals to equal the mean $E(h1_S)$ of the modulated intervals. The denominator of the function shown in Eqn. 4 is slightly different to that shown by Wiegrebe (1998), so that any value of k will produce an I/O function in the range of 0 to 1. Pilot testing revealed that a value of k=1 was sufficient to make the pitch strength of target and non-target intervals indistinguishable when the modulation rate was above the detection threshold. The ideal $h1_S$ of a signal generated using a sinusoidally modulated g is shown before and after being subjected to the expansive function (Eqn. 4.) in Fig. 3.

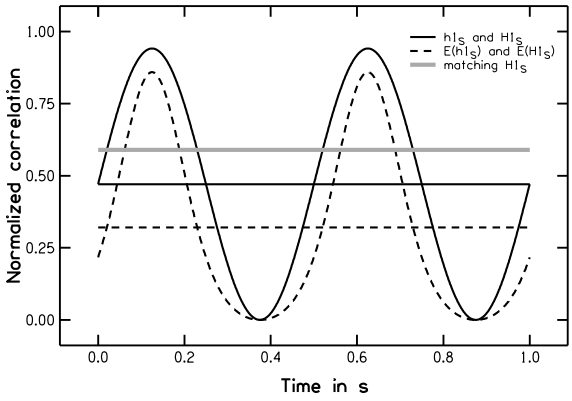

FIG. 3. Diagram illustrating sinusoidally modulated $h1_S$ and the expanded version, $E(h1_S)$. The solid black, horizontal line represents the mean of $h1_S$ (i.e. $H1_S$), and the dashed black line represents the mean of the respective $E(h1_S)$. This is not equal to $E(H1_S)$ because the expansive process is nonlinear. The bold grey line represents the $H1_S$ of the non-target interval IRN required to match the overall perceptual pitch strength of the modulated stimulus.

An adaptive staircase technique was used to measure thresholds where the adaptive parameter was the depth of the gap, or modulation index, depending on the experiment. At the beginning of each threshold run, the gap depth or modulation index of the respective dynamic gain function was set to 0 dB. This was well above the anticipated detection threshold for all stimulus conditions. The adaptive parameter was decreased after two consecutive correct responses and increased after each incorrect response to track the signal level that yielded 70.7% correct responses (Levitt, 1971). A 3-interval task was used because a 3-interval, 3-alternative forced-choice (3I3AFC) task with a 2-down, 1-up rule converges more efficiently than a 2I2AFC task with a 3-down, 1-up rule (Kollmeier et al., 1988). The step size for the increments and decrements in gap depth or modulation index determined by g was 5 dB for the first reversal in level, 3 dB for the second reversal, and 2 dB for the rest of the eight reversals that made up each threshold run. The last six reversals were averaged to obtain a threshold estimate for each run. Three threshold runs were conducted for each participant per stimulus condition using a counter-balanced design to eliminate training effects.

C. Listeners

A total of 8 listeners participated in the current experiments. One subset of 4 listeners (2 male and 2 female, aged between 24 and 27 years) participated in the gap-detection experiment, and the other subset of 4 listeners (2 male and 2 female, aged between 25 and 30 years), one of whom was the author, participated in the modulation-detection experiment. Participants were paid for their services at an hourly rate. Participants had absolute thresholds within 25-dB HL at audiometric frequencies and had

no history of hearing or neurological disorders. The experimental procedures were approved by the Ethics Committee of the University of Nottingham School of Psychology.

IV. RESULTS

A. Measurements

Results from the modulation-detection experiment are shown in Fig. 4. The left-hand panel (A) shows data plotted in the same format as the original intensity TMTF measurements (Viemeister, 1979), where the abscissa is the modulation rate (R_{mod}) in Hz and the ordinate is the modulation index threshold. When plotted in this way, TMTFs resemble the transfer function of a lowpass filter, describing the filtering effect of the integration of the system as a whole. Asterisks adjacent to some of the data points at higher modulation rates represent the number of listeners who were unable to obtain a threshold in those conditions. At the highest modulation rates, some listeners were unable to discriminate the modulated IRN from the unmodulated IRN, even when the modulation depth was 100% (0 dB). This was evidence that the pitch-strength compensation scheme used (Eqn. 4.) successfully prevented listeners from making judgements based on pitch strength alone. The apparent asymptotes observed in the TMTFs towards higher modulation frequencies are artefactual. This was due to a combination of ceiling effects and biasing towards the better performing listeners who were able to obtain thresholds at these rates.

Results from the gap-detection experiment are shown in Fig. 5. The left-hand panel (A) shows data plotted in the same format as the TMTF measurements, allowing for easy comparison. In Fig. 5(A), the abscissa shows the gap rate (R_{gap}), which is the inverse of the gap duration, and the ordinate shows the gap-depth threshold. In contrast with the TMTF results, all listeners were able to obtain thresholds in all conditions measured in the

gap-detection experiment. Gap-detection thresholds decrease as the duration of the gaps increase (i.e. the gap rate decreases). Asymptotic performance was only reached in the 37.50-Hz IRN for the very longest gap duration measured. However, no asymptote in performance was observed for any of the other IRN rates. This could suggest that integration times responsible for limiting resolution are only limited by the stimulus duration. Alternatively, the long gap durations may have invoked a change in listening strategy. When the gap duration approaches the stimulus duration, the task is more closely related to pitch-strength discrimination, as opposed to gap detection. Therefore, the listener was trying to distinguish the stimulus interval with overall weaker pitch, as opposed to listening for the gap within a given stimulus interval. This strategy may have involved use of long-term integration mechanisms like those described in pitch-integration studies (see introduction). Use of this alternative strategy was prevented in the modulation-detection experiments by equalizing the mean pitch strength of stimulus intervals within a given trial.

The pitch-strength TMTFs share the lowpass-filter characteristic observed in intensity TMTFs. The two TMTFs that had low-rate IRN carriers exhibit a band-pass characteristic. This has also been observed in intensity TMTFs at very slow modulation rates and has been partly attributed to a reduction in the number of looks at the envelope fluctuations (Sheft and Yost, 1990, Viemeister, 1979). Similarly, in the data presented here, very few cycles of the modulation were presented to the listeners at the slowest IRN rates, due to the finite duration of the stimuli.

The functions in Figs. 5(A) and 6(A) are different for each IRN rate. The functions shift towards higher modulation rates at higher IRN rates, suggesting that the integration time constants scale with pitch value. The right-hand panels, Figs. 5(B) and 6(B), show the same data but where the abscissa shows the period of the modulation (T_{mod}) or gap (T_{gap})

normalized by d. When plotted in this format, both TMTFs and gap thresholds seem to scale to a single function, with the exception of the highpass regions observed in the lower-rate TMTFs. This indicates that the neural time constants of pitch perception scale linearly with d, as originally suggested by Wiegrebe (2001).

The mean TMTF, excluding TMTFs exhibiting highpass characteristics, cross the 3-dB down point at $T_{mod}/d = 34.2$ (measured from the mean threshold at $T_{mod}/d = 96$ by linear interpolation between neighbouring points). By dividing this value by 2π, the time constant of the system as a whole can be coarsely estimated as 5.44d. Therefore, the time constant is defined as d multiplied by a proportionality constant, described by the symbol, η, from here on. The time constants predicted from the pitch TMTFs ($\eta/150$ = 36 ms, $\eta/37.5$ = 145 ms) are very large in comparison to those of just a few ms derived from TMTFs measured in response to modulations in intensity (Forrest and Green, 1987, Viemeister, 1979). However, the slope of the pitch-domain TMTFs (assessed through linear regression for $T_{mod}/d = 12, 24, 48$) amounted to approximately 4 dB per octave, which was very similar to the roll-off observed in the intensity-domain TMTFs. This suggests that while the time constants of the integration windows may differ markedly between domains, the function of the underlying integration processes may be similar.

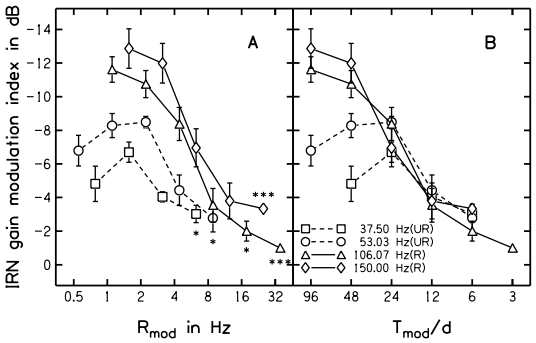

FIG. 4. Both panels show the same data, where each data point is the mean threshold across listeners and error bars represent the inter-listener standard error. In the left-hand panel (A), the ordinate is plotted as the modulation rate in Hz. Asterisks adjacent to data points represent the number of listeners unable to obtain a threshold for those conditions. In the right-hand panel (B), the ordinate is shown as the period of the modulation normalized by the IRN delay. This highlights the scaling of gap-detection threshold with pitch value. Ordinates are reversed to give the TMTF a lowpass-filter shape.

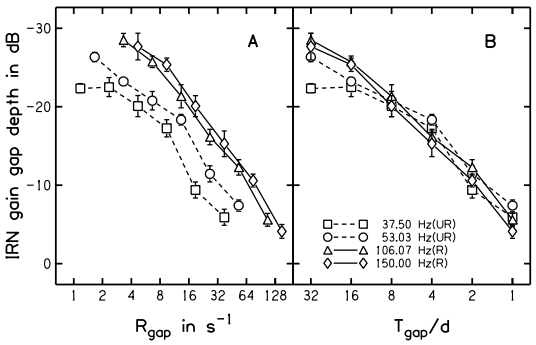

FIG. 5. Both panels show the same data; however, in the left-hand panel, the ordinate is plotted as the gap rate (R_{gap}) in Hz. In the right-hand panel, the ordinate is shown as the period of the gap normalized by the IRN delay. This highlights the scaling of gap-detection threshold with d. Axes are reversed so that the figure is the same format as Fig. 5.

B. Statistical analysis

The statistical significance of the results discussed above was tested by performing linear mixed-models analyses on both modulation and gap data. For the modulation-detection task, the analysis was performed on factors T_{mod}/d and IRN rate. For the gap-detection task, the analysis was performed on factors T_{gap}/d and IRN rate. The dependent variable was mean threshold for each participant in each condition. There was a significant main effect of T_{mod}/d in the modulation experiment [$F(4,48.048)=68.141$, $p<0.001$] and of T_{gap}/d in the gap experiment [$F(5,84.153)=178.603$, $p<0.001$]. The main effect of IRN rate was also significant in both modulation [$F(3,48.038)=7.259$, $p<0.001$] and gap [$F(3,85.007)=178.603$, $p=0.022$] experiments. The interaction of T_{mod}/d and IRN rate was also significant in the modulation-detection experiment [$F(11,48.027)=7.828$, $p<0.001$] but not in the gap-detection experiment [$F(15,69.136)=1.723$, $p<0.066$].

Pairwise comparisons of IRN rate in the modulation-detection experiment show that with the exception of the differences between the 106.07- and 150.00-Hz rates [$F(3,48.037)=14.484$, $p<0.001$], thresholds at all other rate differences were significantly different from one another. The significant differences most likely stemmed from the highpass regions in the TMTFs of the lower-rate IRNs. To test this, pairwise comparisons from the interaction were tested. At T_{mod}/d =6, 12, and 24, thresholds were statistically indifferent. At T_{mod}/d =48, thresholds between the higher-rate 106.07- and 150.00-Hz IRNs were not significantly different [$F(3,48.007)=22.175$, $p=0.199$]; however 37.5- and 53.03-Hz IRN thresholds were different from each other [$F(3,48.007)=22.175$, $p=0.001$] and were both different to thresholds at IRN rates of 150 and 106 Hz [$F(3,48.007)=22.175$, $p<0.013$]. At T_{mod}/d = 96, 106.07- and 150.00-Hz IRN thresholds were not significantly different [$F(3,48.007)=22.798$, $p=0.194$]; however, thresholds for 53.03 Hz were different from both 150 and 106 Hz [$F(2,48.007)=22.798$, $p<0.001$]. There was no data for IRN

rate=37.50 Hz when $T_{mod}/d = 96$ to make a comparison. Taken together, results from the pairwise comparisons confirmed that TMTFs are statistically indifferent when scaled according to d, with the exception of outlying data points of the highpass regions of the lower-rate IRNs.

Pairwise comparisons for each T_{gap}/d in the gap-detection experiment showed that thresholds for each were significantly different from each other at the 0.005 level [$F(5,69.136)=201.633$, $p<0.001$] for all comparisons, with the exception of thresholds between $T_{gap}/d=16$ and 32, where thresholds were still significantly different at the 0.05 level [$F(5,69.136)=201.633$, $p=0.013$]. Pairwise comparisons of IRN rate in the gap-detection experiment showed significant differences between 37.50- and 53.03-Hz IRN thresholds [$F(3,69.919)=3.795$, $p=0.011$] and between 37.50- and 106.07-Hz IRN thresholds [$F(3,69.919)=3.795$, $p=0.002$]. These differences may be due to the fact that the 37.50-Hz IRN was the only rate for which thresholds appeared to reach asymptote by $T_{gap}/d=32$. To test this, pairwise comparisons between thresholds at different IRN rates at each T_{gap}/d were performed. No significant differences were observed between thresholds for the IRNs of different rates at $T_{gap}/d=1,2,4$, or 8; however, thresholds for the 37.50-Hz IRN were different to thresholds for the 106-Hz IRN at $T_{gap}/d=16$ [$F(3,69.305)=2.097$, $p=0.044$], and thresholds for the 37.50-Hz IRN were different to all others at $T_{gap}/d=32$ [$F(3,69.305)=6.125$, $p=0.014$]. Interestingly, the most resolved and most unresolved IRNs had statistically indifferent thresholds overall. Taken together, this suggests that the 37.50-Hz condition was different overall because of the asymptote at long absolute-gap durations.

Combining the observations taken from the statistical analysis of both gap and TMTF data, the post-hoc tests reveal that any main effects of IRN rate were due to measurement-related procedural issues at low gap and modulation rates, not because the

auditory system is using a different processing strategy for the lower-rate IRNs. The harmonic resolvability of the stimuli was determined by the IRN rate, where the two lower-rate IRNs were completely unresolved, whereas the two higher-rate IRNs contained resolved harmonics. Therefore, one can imply that there was no effect of harmonic resolvability, suggesting that pitch was extracted by a temporal mechanism alone, or that spectral and temporal pitch extraction mechanisms feed into a common integrator.

V. MODELLING

A. Rationale

While η was estimated from the TMTFs, the estimate may have been somewhat inaccurate because of the limited number of data points from which it was derived. Furthermore, time constants could not be derived directly from the gap-threshold patterns, as they could be from the TMTFs. Therefore, it was uncertain as to whether the time constants of pitch perception were task-dependent. Use of an auditory model allowed testing of whether a single value of η could accurately predict results from both experimental paradigms.

B. Methods

The first stage of the model consisted of a broad bandpass-filter to simulate the frequency transfer of the outer and middle ear. This filter was a second-order Butterworth filter with a passband between 0.45 and 8 kHz. To simulate the frequency decomposition performed by the cochlea; the signal was multi-band filtered using a 30-channel gammatone filter bank with frequencies evenly distributed on the ERB scale, with frequencies between 0.2 and 8 kHz. To simulate the mechanical-to-neural transduction performed by the inner hair cells and peripheral compression, the signal from each output

of the gammatone filter bank was half-wave rectified and compressed using a logarithmic compression scheme. The signal was subsequently lowpass filtered to simulate the phase-locking limitation of the inner hair cells. This was implemented as a moving-average filter, where the integration window was a 2^{nd}- order exponential function, the time constant of which was set to give a frequency cutoff of 1.2 kHz. This is identical to the default implementation used in the current version of the AIM software package (aim2009[1]). The resulting multi-channel signal is referred to as the neural activity pattern (NAP). The NAP is a per-channel probability of neural firing over time. The decision statistic was derived from the instantaneous temporal regularity within the NAP, $h1_{NAP}$. This was generated by firstly taking the cross product of the NAP at the time lag equalling the IRN delay. Information about the level of the stimulus was removed by normalizing the cross product by the mean power of the NAP across channels and time. Normalized cross products were generated for 1000 stimuli based on unique noise sources, which were then averaged to reduce the stimulus-induced noise. This multi-channel internal representation was then averaged across channels and convolved with an exponentially decaying window, resulting in an internal representation of the running autocorrelation, $R(h1_{NAP})$. The values that an exponentially decaying window returns are negligible beyond 3 times its time constant, and so integration windows were limited to this length for computational efficiency. The beginning of $R(h1_{NAP})$ was truncated by an amount equal to the duration of the integration window. This was to remove the initial build-up from 0 to a stable level. The decision statistic, D, was then calculated as the maximum of $R(h1_{NAP})$ over time minus the minimum. To sample the range of listener thresholds, gap stimuli were generated with gap depths ranging from -32 dB to 0 dB in 8-dB steps. Modulated stimuli

[1] Available from http://www.pdn.cam.ac.uk/groups/cnbh/research/aim.php

were created with modulation indices ranging from -20 to 0 dB in 5-dB steps. Gap stimuli were generated for all conditions measured in the gap experiment. However, modulated stimuli were only generated for the pair of IRNs with higher rates to omit the lower-rate IRNs where the TMTFs exhibited highpass behaviour at low-frequency modulations, that the model presented here was not designed to account for. For each experimental condition, D was calculated as a function of either gap depth, or modulation index, depending on the stimulus type. Threshold was defined as the modulation depth at which D reached a criterion level, C, and this criterion was the main parameter in the fitting process. Both gap and modulation thresholds were fitted simultaneously, with a fixed value of C, and C was then varied to find the value that minimized the root-mean-squared (RMS) deviation between the simulated and observed thresholds. This fitting process was repeated for integer η values in the range of 1 to 11.

C. Predictions

Fig. 6 shows the predicted thresholds superimposed upon the listener data. The abscissa is different to those used in the figures presented in the results section (Figs. 5 and 6). By representing both modulation and gap rates in terms of $\log_2(d/ T_{gap})$ and $\log_2(d/ T_{mod})$, data from both gap- and modulation-detection tasks can be presented on the same set of axes. Each panel shows the predictions for a different value of η, illustrating the systematic effect of increasing η. When η was small, the time constants were relatively short, and therefore simulated thresholds did not begin to roll off until relatively fast gap or modulation rates. As η was increased, the time constants became longer, smearing larger features in the $h1_{NAP}$, and so simulated thresholds began to roll off at low gap and modulation rates.

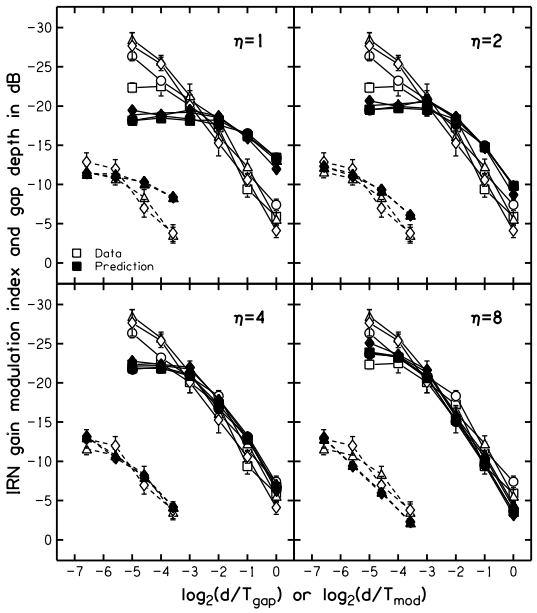

FIG. 6. Each panel shows the model predictions overlaid on top of the data recorded in the experiment for different values of η, denoted in the top right of each panel. Each panel shows both TMTF and gap-detection data scaled by the IRN rate. The ordinate is plotted as a $\log_2$ scale so that both gap and TMTF thresholds can be plotted clearly on the same axes. As η increases from 1 to 8, the simulated cutoff shifts rightwards.

Fig. 7(A) shows how the RMS error of the simulation varied with η when TMTF and gap-detection data were fitted simultaneously. The best overall fit was achieved when η was 7. Fig. 6 shows that η=8 also produced a reasonably good fit to the listener data overall. However, on careful inspection of the modelled thresholds in Fig. 6, it can be seen that, in general, gap thresholds are more accurately simulated using higher values of η, and the cutoff of the TMTFs are more accurately simulated using lower values of η. If TMTF and gap thresholds are modelled independently, then the best-fitting values of η are 4 and 7 respectively, as shown in Fig. 7(B) and (C).

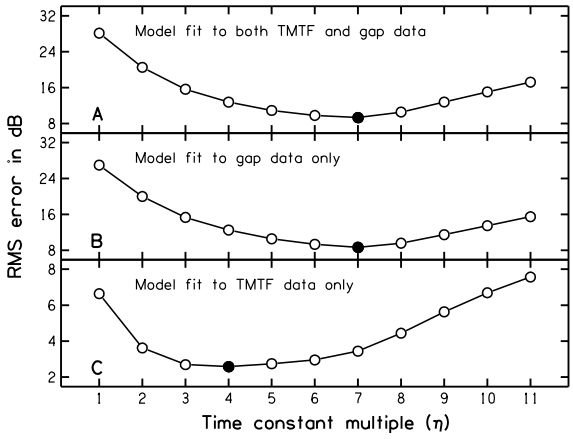

FIG. 7. RMS error of the fitting process as a function of η for (A) TMTF and gap data simultaneously fit, (B) model fit to gap data only, and (C) model fit to TMTF data only. The filled symbol in each panel shows the lowest error point.

VI. DISCUSSION

In the current study, the temporal resolution of the monaural pitch-extraction mechanism was measured using pitch-domain analogues of standardised intensity envelope resolution paradigms. The TMTF measurements were able to separately quantify the temporal smoothing imposed by integration from the sensitivity of the system to the modulations, thereby providing compelling evidence that the time constants of neural pitch extraction scale with the interval of temporal regularity within the stimulus. The time constants derived directly from the TMTFs scaled with the stimulus rate by a factor η =5.44. An auditory model was used so that data from gap- and modulation-detection tasks could be compared. For the TMTF data, a value of η=4 was found to be most appropriate to describe the neural integration time constants. A slightly higher value of η=7 was required to minimize the RMS deviation between simulated and measured gap-detection data. However, this was unlikely to reflect a task-dependent difference in the integration time constants, because when the gap duration approached the stimulus duration, listeners

probably changed listening strategy from gap-detection to pitch-strength discrimination. This meant that no asymptote in thresholds was observed towards longer gap durations. Use of longer time constants in the model shifted the predicted asymptote towards longer gap durations, thus reducing the RMS deviation between simulated and measured thresholds. The model was able to accurately predict the sensitivity difference between gap and modulation detection thresholds when the two were fitted using the same criterion, thus providing good evidence that the time constants responsible for limiting resolution did not vary according to the task.

Interestingly, no effects of harmonic resolvability were observed. This was surprising in the context of results from studies in which the effect of stimulus duration on pitch-value discrimination thresholds were measured (Plack and Carlyon, 1995, White and Plack, 1998). These studies generally showed large differences between the stimulus duration required for discrimination performance of resolved and unresolved stimuli to reach asymptote, suggesting that different pitch-extraction mechanisms were associated with each. However, a resolution task was used in the current study; the effects observed using a pitch-value discrimination integration paradigm may well be based on different pitch-extraction mechanisms. While the limit of human phase locking is not known, the frequency range in which stimuli were presented in the current study would be expected to contain at least some frequencies within the putative phase-locking limit. Therefore, results from the current study suggest that if reliable TFS is available, the initial pitch-extraction process presumed responsible for limiting temporal resolution is based on a temporal mechanism, or at least spectral and temporal mechanisms that feed into a common integration process, or a pair of identical integration processes.

Another major finding of the current study was that the pitch-extraction mechanism was equally sensitive to changes in temporal regularity, irrespective of the

repetition rate of the stimulus. Pressnitzer et al. (2001) showed that, in order to account for the lower limit of melodic pitch in a temporal model of pitch perception, a weighting function could be applied that progressively reduced the output of the time-interval histogram towards longer lags. In an unrelated study, Krumbholz et al.(2003a) measured the detectability of a tonal signal in the presence of a tonal masker. To account for the experimental results of this study using a temporal pitch model, weighting functions similar to that suggested by Pressnitzer et al. (2001) were applied to the simulated time-interval histograms. However, the asymptotic thresholds of the TMTFs measured in the current study were equal (for those not exhibiting the band-pass characteristic), irrespective of the IRN rate. This suggests that the system is equally sensitive to modulations in temporal regularity, irrespective of the stimulus rate. This finding is not compatible with models that apply time-interval weightings to reduce the pitch-value resolution towards the lower limit of pitch, as the weighting would also reduce the sensitivity of the model to modulations in pitch strength at lower repetition rates. An alternative theory is that the widths of the bins that comprise the internal time-interval histogram are not equal, but greater, at longer time intervals. This alternative model would lower the frequency resolution towards the lower limit of pitch, but maintain the model's sensitivity to modulations in temporal regularity.

The temporal resolution of the monaural auditory system measured in the current study exhibits some striking similarities to the temporal resolution observed in binaural processing. The term "binaural sluggishness" is commonly used to refer to the inability of the binaural system to follow fast changes in interaural parameters over time when compared to the exquisite temporal resolution of the monaural auditory system in response to changes in intensity. The sluggishness observed in binaural processing is thought to reflect the relatively long integration window, estimated to be in the range of several tens

to a few hundred milliseconds, depending on the experimental conditions (Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998; Akeroyd and Summerfield, 1999). Similarly, the value of $\eta = 4$ used to simulate the TMTF measurements in the current study implies pitch-integration time constants in the range of 27 ms (for the 150.00-Hz IRN) to 107 ms (for the 37.50-Hz IRN). The sluggish response of both binaural and pitch mechanisms may reflect the similarities in the underlying processing mechanisms, in that both pitch and binaural information may be extracted using analogous, correlation-based mechanisms. Therefore, it is possible that the time constants associated with binaural processing may scale according to the interval of interaural temporal regularity (interaural time difference) in a binaural signal, just as pitch-processing time constants appear to scale according to the interval of temporal regularity within a monaural stimulus. The binaural system processes interaural time differences in the range of only a few tens of microseconds, whereas the pitch processor works in the order of milliseconds; therefore, one would expect the binaural value of η to greatly exceed its monaural pitch counterpart. This hypothesis has yet to be tested; however, it would be relatively simple to replicate the current study in the binaural domain.

Pitch is known to be one of the most important cues for helping listeners to hear out speech in noisy backgrounds, particularly in backgrounds of competing speech. Speech signals vary rapidly over time; therefore, one would expect the sluggishness of the pitch-extraction mechanism to be a hindrance when trying to follow the pitch-related changes in voiced speech. However, in an integration study, Plack and White (2000b) have shown that gaps in the intensity of tonal stimuli of as little as 4 to 8 ms were able to reset the pitch-integration mechanism. In a later study (Plack and White, 2000a), they showed that pitch information was integrated across gaps of 8 and 16 ms between tone bursts when the gaps were filled with noises with similar energy spectra to the tonal

portions of the stimulus. Therefore, it is possible that the time constants that limit temporal resolution are also resettable, depending on changes in the stimulus intensity. Based on Plack and White's (2000) findings, task-dependent differences may not have been observed in the current study because the stimuli had relatively constant energy spectra over time. However, the intensity fluctuations in running speech may reset the integration window, based on top-down feedback mechanisms such as those proposed in the model of Balaguer-Ballester et al. (2009), thereby improving the temporal resolution of the monaural pitch-extraction mechanism.

Chapter 2

The temporal resolution of pitch perception II: Effects of frequency region

I. INTRODUCTION

In Chapter 1, a novel stimulus based on iterated rippled noise (IRN) was presented that allowed measurement of pitch-domain analogues of temporal modulation transfer functions (TMTFs) and gap-detection thresholds. The results from that study suggested that the time constants of the leaky-integration window presumed responsible for limiting temporal resolution scale according to the repetition rate of the stimulus, while sensitivity to modulations in pitch strength is independent of IRN rate. The harmonic resolvability of the stimuli was included as a parameter, but, surprisingly, no effects of resolvability were observed on either the sensitivity of the system or the scaling of integration time constants.

The integration of pitch information has also been measured in tasks that quantify the ability of the auditory system to combine information across time in order to improve performance in pitch discrimination. In these pitch-integration studies, it is assumed that discrimination thresholds will decrease with increasing stimulus duration until the system has reached its integration capacity. Once the integration window has been filled, longer stimulus durations provide no performance benefits. Results from some of these studies (Plack and Carlyon, 1995, White and Plack, 1998) suggest that the duration of the integration windows is dependent on the harmonic resolvability of the stimuli.

The discrepancies between data measured in integration and resolution studies suggest that the effects of stimulus harmonic resolvability may be dependent on the task. Functional models designed to simultaneously account for behavioural data from both integration and resolution tasks generally consist of two separate integration processes: a lower-level short-term integration process to explain resolution data, and a higher-level and longer-term integration process to explain integration data. Should such an arrangement of mechanisms exist in the auditory system, it is possible that pitch would be extracted differently by each mechanism. Alternatively, the effects of harmonic

resolvability may be dependent on the listening region in which the stimuli are presented. Acoustic waveforms generally consist of a rapidly fluctuating carrier signal that is modulated by a slowly varying intensity envelope. There is a phase-locking limit to which the mechanical-to-neural transduction process is able to transmit the timing of peaks in the temporal fine structure (TFS) to the central auditory system. For humans, the breakdown of phase locking is often modelled as a lowpass filter with a cutoff of 1.2 kHz. In a listening region below the phase-locking cutoff, high-fidelity TFS information is available to convey the frequencies of the resolved harmonic components of complex tonal sounds. In the frequency region above the phase-locking cutoff, the transmission of the TFS from each harmonic of a resolved stimulus would be severely degraded. In contrast, the relatively slow within-channel interactions between unresolved harmonics would still be transmitted accurately. These harmonic interactions have the same periodicity as the periodicity of the stimulus waveform. Therefore, in higher frequency regions where TFS is degraded, it would be more likely that a spectral mechanism would extract pitch from resolved stimuli, whereas a temporal mechanism would be expected to extract pitch from unresolved stimuli.

In the resolution study conducted in the previous chapter, comparisons between resolved and unresolved stimuli were made within a relatively low spectral region (0.78 to 2.98 kHz) where high-fidelity TFS would have presumably been available. However, in the integration study of Plack and Carlyon (1995) where an effect of harmonic resolvability was observed, a 62.5-Hz unresolved stimulus and a 250-Hz resolved stimulus were presented within a relatively high-frequency band between 1.38 and 1.88 kHz; a 250-Hz unresolved stimulus band limited to an even higher region between 5.50 and 7.50 kHz, and a resolved 250-Hz stimulus lowpass-filtered below 1.88 kHz were also used. Thus, only the resolved 250-Hz stimulus was presented with spectral energy below 1.2 kHz. This

means that no direct comparisons of resolved and unresolved stimuli were made in a frequency band containing high-fidelity TFS information. Similarly, in the study of White and Plack (1998) where an effect of harmonic resolvability was also observed, a 250-Hz resolved stimulus was presented lowpass-filtered below 1.88 kHz, and a 62.5-Hz resolved stimulus was presented lowpass-filtered below 0.47 kHz. Additionally, a 250-Hz unresolved stimulus band-limited between 5.50 and 7.50 kHz, and a 62.5-Hz unresolved stimulus band-limited between 1.38 and 1.88 kHz were used. Again, no direct comparisons of resolved and unresolved stimuli were made in a frequency band containing high-fidelity TFS information. In White and Plack (2003), integration times were found to scale with the stimulus rate, but only unresolved stimuli in frequency regions (2.75 – 3.75 and 5.50 – 7.50 kHz) well above the phase-locking limit were used. In the study of Krumbholz et al. (2003), where integration times were also shown to scale with the stimulus rate, 31.25-, 62.5-, 125-, and 250-Hz stimuli were presented band-limited between 800 and 3200 Hz. In this spectral band, the 250-Hz stimulus would have contained resolvable components, whereas the 32.5-Hz stimulus, at the other extreme, would not. Interestingly, no effects of resolvability were observed, perhaps because the listening region used was very similar to that described in Chapter 1, which also found no effects of resolvability.

The aim of the current study was to test whether the effect of resolvability is dependent on the task of integration, as opposed to resolution, or on the listening region in which the stimuli are presented. This was achieved by repeating the experiment presented in Chapter 1 in a listening region above the putative phase-locking limit. While the exact frequency at which phase locking deteriorates in humans is not known, the fidelity of the TFS information in a higher-frequency listening region would be expected to be degraded,

compared to the fidelity of the TFS information in the lower-frequency listening region used in Chapter 1.

II. EXPERIMENT 1: MEASUREMENT OF THE TEMPORAL RESOLUTION OF PITCH PERCEPTION IN A HIGH-FREQUENCY REGION

A. Methods

1. Stimuli

The temporal resolution of pitch perception was measured in a high-frequency listening region using the gap-detection and TMTF paradigms described in Chapter 1. In Chapter 1, stimuli were filtered between 0.78 and 2.98 kHz using IRN rates of 37.50, 53.03, 106.07, and 150.00 Hz. In that study, the two lower IRN rates were unresolved and the two higher IRN rates contained resolved harmonics. The lower (unresolved) IRN rates used in the current study were set to coincide with the higher (resolved) IRN rates used in the previous study to disambiguate between potential effects of harmonic resolvability, listening region, and IRN rate. For this, thresholds were measured for IRN repetition rates that were 1 octave above (424.26 Hz) and below (106.07 Hz) and 0.5 octaves above (300.00 Hz) and below (150.00 Hz) a central value of 212.13 Hz (212.13 Hz is 0.5 octaves above 150 Hz). To test the effects of harmonic resolvability, stimuli were filtered into a 2.2-kHz bandwidth with a centre frequency of 3.74 kHz. The lower cutoff frequency of the spectral band was set at 2.64 KHz, which coincides with the mean value of harmonics (2.625) per 10-dB auditory filter bandwidth, in order to achiever partial resolvability at a repetition rate of 212.13 Hz. As for the low-frequency band used in Chapter 1, the 2 lower repetition rates (106.07 Hz, 150.00 Hz) were completely unresolved, whereas the 2 higher rates (300.00 Hz, 424.26 Hz) contained resolved components. To aid the description

given, the connection between IRN rates, listening regions, and harmonic resolvability between both Chapters 1 and 2 is shown graphically in Fig. 1.

As in Chapter 1, stimuli were presented at a level of 65-dB sound pressure level (SPL) and were gated on and off with 5-ms cosine-squared ramps. Stimuli were presented in a continuous noise to mask audible distortion products below the stimulus passband using the same methods and equipment described in Chapter 1.

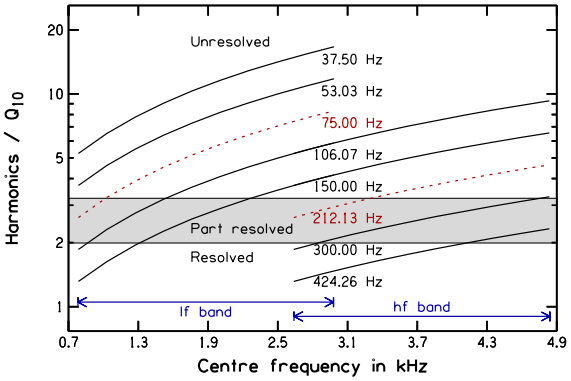

FIG. 1. Graphical representation of the parameter space showing the relationship between the IRN rates used in Chapters 1 and 2. The abscissa represents the centre frequency of the auditory filters across the listening regions in which the stimuli were presented, and the ordinate represents the number of harmonics of the IRN spectra falling into the 10-dB bandwidth of those filters. The blue arrows mark the low- and high-frequency regions in which stimuli were presented. The parameter is the rate of the stimuli, where the black solid lines represent the 4 IRN rates used in each spectral band. The IRN rates are given by the text below each curve. Note the overlap of the 106- and 150-Hz conditions between bands. The shaded area in the centre of the figure shows the region of partial harmonic resolvability according to the rule of Shackleton and Carlyon (1994). The dashed red lines correspond to the limit of harmonic resolvability at the lower edge of each band (2.625 harmonics/Q_{10}). The higher-rate IRNs within each band contain

some resolved harmonics, while the lower-rate IRNs are completely unresolved throughout each band. The high-frequency band is the subject of this study.

2. Procedure

For IRN rates of 106.07 and 150.00 Hz, gap-depth detection thresholds were measured for gap durations, T_{gap}, equal to multiples of 2, 4, 8, 16, 32, and 64 times the IRN delay, d. The modulation-detection experiment was more difficult, requiring slower modulation rates to achieve a good dynamic range of thresholds. Therefore, modulation-detection thresholds were measured for modulation periods (T_{mod}) equal to multiples of 6, 12, 24, 48, 96, 192, and 384 times d. The longest T_{mod} used for the 106.07- and 150.00-Hz IRNs were limited to 192 times the respective d, as use of a T_{mod} = 384d would have required a stimulus duration in excess of 3 seconds to capture a single modulation cycle. The stimulus durations were the same as those used in the previous chapter: 1.2068 seconds for the gap-detection experiment and 1.8102 seconds for the modulation-detection experiment. This allowed for at least one complete modulation cycle at the slowest modulation rate used in the experiment. Thresholds were measured using the same adaptive staircase procedure as in Chapter 1.

3. Listeners

The same group of 8 listeners who participated in Chapter 1's experiments participated in the current experiments. The same subset of 4 listeners (2 male and 2 female, aged between 24 and 27 years) participated in the gap-detection experiment, and the other subset of 4 listeners (2 male and 2 female, aged between 25 and 30 years) participated in the modulation-detection experiment, one of whom was the author; the others were paid for their services at an hourly rate.

B. Results

1. Measurements and interim discussion

Thresholds measured in the current study in the high-frequency band (2.64 - 4.84 kHz) are shown in Fig. 2. TMTFs for the lower-rate IRNs in the high-frequency region exhibited a similar bandpass characteristic as for the lower-rate IRNs in the low-frequency region (measured in Chapter 1), because the periodicity of the lowest modulation rates used approached the duration of the stimulus (for a detailed discussion, refer to Chapter 1). As in Chapter 1, the asymptotic thresholds in the high-frequency region TMTFs that did not exhibit the bandpass characteristic were equal for IRNs of different rate, which again suggests that sensitivity is not dependent on IRN rate.

In Chapter 1, both TMTF and gap thresholds scaled to a single function when gap and modulation rates were normalized by d. The same also happened for thresholds measured in the current study (right-hand panels of Figs. 2 and 3), suggesting that the time constants of pitch perception also scale with pitch value in the high-frequency band. Importantly, as for the low-frequency region, the scaling of thresholds was independent of the harmonic resolvability of the stimuli, suggesting that the pitch of both resolved and unresolved stimuli are extracted using a common integration window in the high-frequency band.

In the current experiment, all listeners were able to obtain thresholds for the lower-rate 106.07- and 150.00-Hz IRNs. However, when the gap duration was 2d, they were unable to perform the task for the 300.00- and 424.26-Hz IRNs. Thus, listeners were unable to detect gaps much smaller than ~10 ms (4/424.26=9.43 ms), suggesting that while time constants generally appear to scale with d, there is an absolute minimum integration time. An absolute minimum neural integration time associated with the pitch-

extraction mechanism has also been suggested by Wiegrebe (2001). Furthermore, the absolute minimum integration time may depend on the spectral band, as listeners were readily able to detect gap durations of 1d for the 150.00-Hz IRN in the low-frequency band (1/150=6.7 ms).

To enable comparison between the low-frequency region data measured in Chapter 1 and the high-frequency region data measured in the current study, thresholds from each study are plotted on the same axes in Fig. 4. Thresholds are only shown for the highest IRN rate used in each band to simplify the comparison. The highest IRN rates were selected, as the associated TMTFs did not exhibit the artefactual bandpass characteristic. Thresholds from each frequency band were clearly different at equal T_{mod}/d and T_{gap}/d. No asymptotes were observed in the scaled gap-detection data; therefore, the differences in thresholds measured between bands could be a vertical separation, suggesting a sensitivity difference, or a horizontal separation, suggesting an integration time difference, or even a combination of the two. However, asymptotes were observed in the scaled TMTF data from each band. The asymptotic thresholds between bands were similar, suggesting a constant sensitivity across bands. Therefore, the differences between data from each band must be due to a difference in integration time constants. The asymptote in the high-frequency band data occurred at a lower T_{mod}/d relative to the asymptote in the low-frequency band data, suggesting that the integration time constants are longer in the high-frequency band. Therefore, the scalar factor (η) that relates the integration time constants to d is dependent on the frequency region in which the stimuli are presented.

The time constants of the system as a whole can be estimated directly from the TMTF data. The high-frequency band TMTF shown in Fig. 4 crossed the 3-dB-down point at approximately $T_{mod}/d = 106.3$ (measured by linear interpolation between neighbouring points from the lowest threshold: $T_{mod}/d = 192$). By dividing this by 2π, the value of η can

be estimated as 16.92 in the high-frequency band. This is over three times larger than the value of η=5.44 derived from the low-frequency band data in Chapter 1. The slopes of the roll-offs associated with the TMTFs in Fig. 4 (assessed through linear regression for the highest 3 modulation rates) both amounted to ~4 dB/octave.

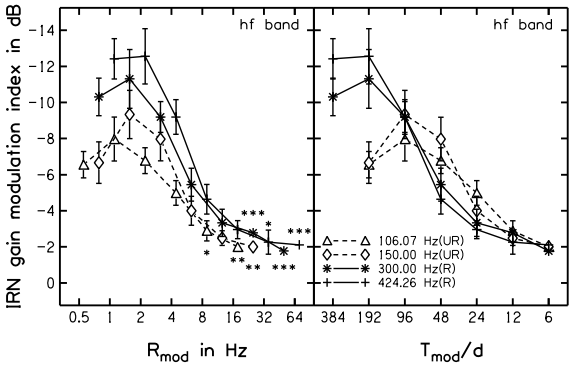

FIG. 2. Mean modulation-detection thresholds averaged across 5 listeners, where error bars represent inter-listener standard error. The left-hand panel shows data plotted in the same format as the original intensity TMTF measurements (Viemeister 1979), where the abscissa is the modulation rate (R_{mod}) in Hz and the ordinate is the modulation index at threshold. Asterisks adjacent to some of the data points represent the number of listeners who were unable to obtain a threshold in the respective conditions, which was generally the case at higher modulation rates. The same data from the left-hand panels is shown in the right-hand panels, where the ordinate is the modulation period normalised by the IRN delay. As in Chapter 1, this shows the scaling of thresholds with pitch value.

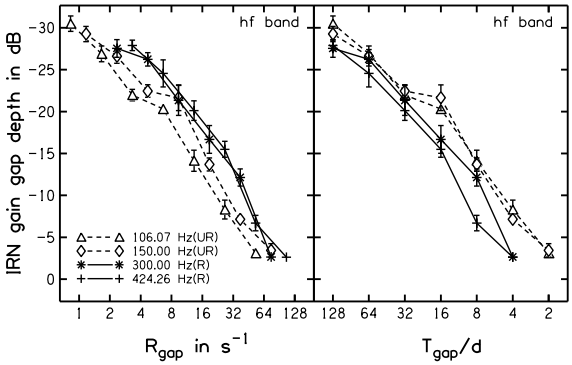

FIG. 3. Mean gap-detection thresholds averaged across 5 listeners, where error bars represent inter-listener standard error. The organisation of the panels is the same as in Fig. 2. The left-hand panels show gap-detection data plotted with reversed axes in a similar format to the TMTF data, allowing for easy comparison of results.

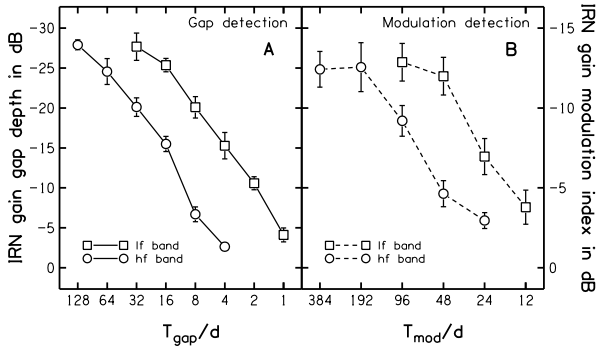

FIG. 4. Thresholds from the highest IRN rate used in each of Chapters 1 and 2, plotted adjacent to one another to show the effect of listening region. Mean gap-detection thresholds are shown in the left panel, and mean modulation-detection thresholds are shown in the right panel.

2. Statistical analysis

The statistical significance of the results discussed above was tested by performing linear mixed-models analyses on both modulation- and gap-detection data in the high-frequency region measured in the current study. See Chapter 1 for results of a similar analysis performed on the low-frequency band data. For the modulation-detection task, the analysis was performed on factors T_{mod}/d and IRN rate. For the gap-detection task, the analysis was performed on factors T_{gap}/d and IRN rate. The dependent variable was mean threshold for each participant in each condition.

There was a significant main effect of T_{mod}/d in the modulation experiment [$F(6,60.054)=70.511$, $p<0.001$] and of T_{gap}/d in the gap experiment [$F(6,75)=510.092$, $p<0.001$]. There were also significant interactions of T_{mod}/d with IRN rate $F(16,60.022)=5.597$, $p<0.001$] and T_{gap}/d with IRN rate [$F(16,75)=2.654$, $p=0.002$]. The main effect of IRN rate was significant for the gap-detection data [$F(3,75)=34.462$, $p<0.001$], but not for the modulation-detection data [$F(3,60.043)=0.966$, $p=0.415$]. Pairwise comparisons of IRN rate in the gap-detection data indicated that the main effect was due to the thresholds for the 424.26-Hz conditions being significantly different to the rest of the IRN rates [$F(3,75)=5.599$, $p=0.002$], while thresholds for all other IRN rates were statistically indifferent. Pairwise comparisons of the interaction between T_{gap}/d and IRN rate show that the 424.26-Hz IRN was only significantly different from all other thresholds when T_{gap}/d = 8 [$F(3,75)=17.282$, $p<0.001$]. Pairwise comparisons of T_{gap}/d were all significantly different from one another, as were all T_{mod}/d, with the exception of the difference between T_{mod}/d = 96 and T_{mod}/d = 192 [$F(6, 60.006)=75.892$, $p=0.474$]. This was likely due to the bandpass characteristic exhibited by the lower-rate IRNs. To test this, pairwise comparisons between IRN rates at T_{mod}/d = 192 were made. At this modulation rate, 106.07- and 150.00-Hz IRN thresholds were not significantly different. The 300- and

424.26-Hz IRN thresholds were not significantly different either, but both of the lower-rate IRN thresholds were significantly different from the higher-rate IRN thresholds [$F(3,60.006)=20.839$, $p<0.001$]. Thresholds between $T_{mod}/d = 96$ and $T_{mod}/d = 192$ were significantly different for both of the higher rate IRNs: 300.00 Hz [$F(6,60.036)=25.890$, $p=0.033$], 424.26 Hz [$F(6,60.028)=40.723$, $p=0.001$]. Thresholds between $T_{mod}/d = 192$ and $T_{mod}/d = 384$ were not significantly different for either of the higher rate IRNs: 300.00 Hz [$F(6,60.036)=25.890$, $p=0.304$], 424.26 Hz [$F(6,60.028)=40.723$, $p=0.886$]. This analysis suggests that TMTFs that did not exhibit the bandpass characteristic had reached asymptote by $T_{mod}/d = 192$.

To investigate the significance of the difference between high- and low-frequency bands, another two linear mixed-models analyses were performed separately for modulation- and gap-detection data across both spectral bands (data from both the current study and from Chapter 1). This revealed a significant main effect of frequency band for both the gap-detection experiments [$F(1,147)=216.177$, $p<0.001$] and the modulation-detection experiments [$F(1,111.017)=52.205$, $p<0.001$].

C. Modelling

1. Time constants of the leaky-integration windows

In the previous chapter, an auditory model of temporal pitch extraction was used to simultaneously predict both gap- and modulation-detection thresholds. The fit produced was reasonably accurate when the time constants used in the leaky-integration process scaled with the IRN delay, d, by a factor, η. Like in the low-frequency band companion study, no significant effect of harmonic resolvability was observed in the current high-frequency band study; therefore, it was presumed that a similar temporal pitch-extraction model would be able to account for the current data. However, η values derived from the

TMTF results were substantially different between the two listening regions, and so in the simulations presented here, independent values of η for each frequency band were free parameters in the fitting process.

The model used was almost identical to that presented in Chapter 1. The first stage of the peripheral model consisted of a broad bandpass filter to simulate the frequency transfer of the outer and middle ear. This filter was a second-order Butterworth filter with a passband between 0.45 and 8 kHz. To simulate the frequency decomposition of the cochlea, the signal was multi-band filtered using a 30-channel gammatone filter bank with frequencies evenly distributed on the ERB scale between 0.2 and 8 kHz. To simulate the mechanical-to-neural transduction performed by the inner hair cells and intensity compression, the signal from each output of the gammatone filter bank was half-wave rectified and compressed using a logarithmic compression scheme. The resulting multi-channel probability of neural firing is referred to as the neural activity pattern (NAP). Normally at this stage, the signal would be lowpass-filtered to simulate the phase-locking limitation of neural transduction. However, it was noticed that the lowpass filter had a dramatic effect on modelling the differences between the low- and high-frequency regions and was thus omitted at first. The implications of a simulated phase-locking limitation are discussed in a separate analysis in the current chapter. The NAP was then used to calculate the simulated internal estimate of instantaneous temporal regularity, $R(h1_{NAP})$, using the methods described in Chapter 1.

The decision statistic, D, was then calculated as the maximum of $R(h1_{NAP})$ minus the minimum of $R(h1_{NAP})$ in response to the stimulus. To simplify the modelling process, thresholds were simulated for the limited data set shown in Fig. 4. To sample the range of listener thresholds, pitch-gap stimuli were generated with gap depths ranging from -32 dB to 0 dB in 8-dB steps. Pitch-modulation stimuli were created with modulation indices

ranging from -20 to 0 dB in 5-dB steps. For each experimental condition, D was calculated as a function of either gap depth or modulation index. Threshold was defined as the SMR at which D reached a criterion level, C, and this criterion was the main parameter in the fitting process. Both gap and modulation thresholds were fitted simultaneously, with a fixed value of C, and C was then varied to find the value that minimized the root-mean-squared (RMS) deviation between the simulated and observed thresholds. This fitting process was repeated using 2 free parameters (η in the low- and high-frequency bands) to find the combination of η that best described the results. A range of η from 1 to 9 was used in the low-frequency band, while a range of η from 1 to 39 was used in the high-frequency band.

The results of the fitting process are shown in Fig. 5. The integer values of η that gave the minimum RMS deviation between listener data and predicted thresholds were 7 in the low-frequency band and 28 in the high-frequency band. These values were slightly larger than those predicted directly from the TMTF measurements. This was because TMTF and gap-detection thresholds were fitted simultaneously and the absence of an asymptote in the gap-detection data forced the model to use longer time constants than would be predicted based on the TMTF data alone to obtain a reasonable fit (for a detailed discussion refer to Chapter 1). The data suggested that the auditory system is equally sensitive to changes in $h1_S$, irrespective of the listening region in which the stimuli are presented. The current analysis demonstrated that when the phase-locking filter was disabled, the model was also equally sensitive to changes in $h1_S$, irrespective of the listening region in which the stimuli were presented. Furthermore, the model was able to predict the higher sensitivity to gaps in $h1_S$ compared to modulations in $h1_S$.

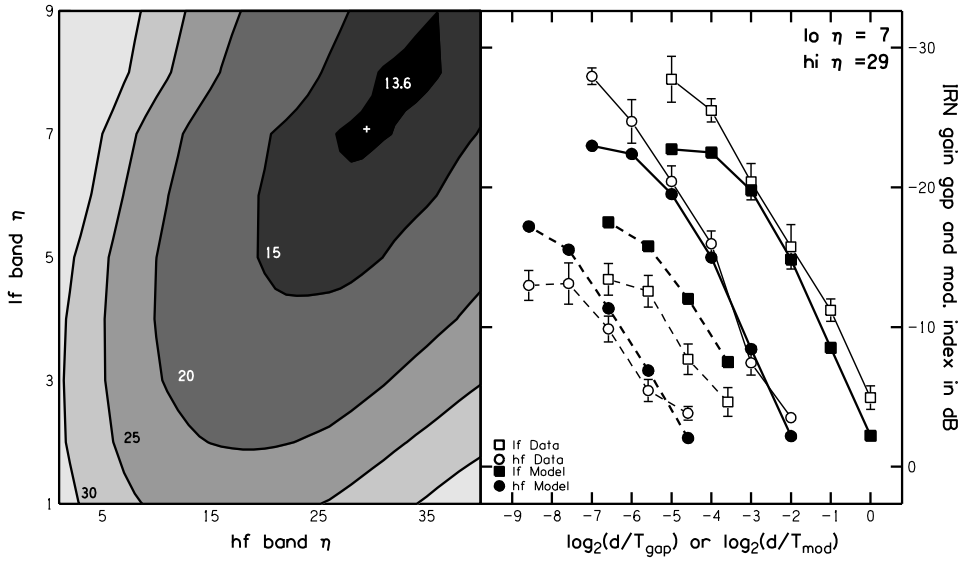

FIG. 5. The left-hand panel shows a contour plot of the RMS deviation between the predicted and measured thresholds of the data shown in Fig. 4. The abscissa shows high-frequency band η, the ordinate shows low-frequency band η, and the shading shows RMS deviation between simulated and measured thresholds, where darker shading represents lower error. Numbers adjacent to contour bands represent the maximum RMS deviation in dB within each *bounded region. The '+' symbol denotes the point of minimum error (low* band $\eta = 7$, high band $\eta = 29$). The model predictions when using the best combination of η are displayed in the right-hand panel (filled symbols), superimposed upon the listener data (open symbols) from Fig. 4. Squares and circles represent low- and high-frequency band data predictions respectively. As in Chapter 1, T_{gap}/d and T_{mod}/d are plotted on a logarithmic scale to enable thresholds from both experiments to be shown on the same axes. The ordinate shows gap-detection thresholds in dB.

2. Implications of a simulated phase-locking limitation

The phase-locking filter was disabled in the simplified model presented in the previous analysis. The aim of the current analysis was to assess the effects of a simulated phase-locking filter on the $h1_{NAP}$ in response to the gap stimuli used in the current study.

For this, the signal within each channel of the NAP was processed by a 2nd order lowpass filter, and $h1_{NAP}$ was then calculated from the processed NAP. The effects of filters with cutoff frequencies of 3.0 kHz and 1.2 kHz were tested and compared to $h1_{NAP}$ when no phase- locking filter was used. Fig. 6 shows the $R(h1_{NAP})$ of gap stimuli with gap depths of 0 dB and gap durations of 4d, presented in both low- and high-frequency listening regions. For demonstrative purposes, a fixed value of $\eta=1$ was used in both frequency regions. This was to ensure that the time constants of the integration windows were small relative to the gap durations and so the effect of the phase-locking filter was not confused with the effect of integration. When the phase-locking filter was disabled, the gap depths in $R(h1_{NAP})$ were equal in stimuli presented in both of the low- and high-frequency bands. When the phase-locking filter had a lenient cutoff of 3.0 kHz, the gap depths in $R(h1_{NAP})$ in stimuli presented in high- and low-frequency regions were both reduced. However, the gap depth in $R(h1_{NAP})$ in the high-frequency region stimulus was reduced more compared to that in the low-frequency region stimulus. When the phase-locking filter had a more realistic cutoff of 1.2 kHz, the gap depth in $R(h1_{NAP})$ in response to the high-frequency listening region stimulus was greatly reduced relative to that of the low-frequency listening region stimulus.

In order to understand why the gap depths were reduced by the introduction of the phase-locking filter, summary autocorrelograms of the NAP in response to regular (no gap) IRNs from the low- and high-frequency bands were compared (Fig. 7). The summary autocorrelograms were generated by subjecting each channel of the NAP to an autocorrelation integrated across the entire stimulus duration, then averaging the resulting autocorrelograms across channels and normalizing the mean autocorrelogram to remove level information. The height of the peaks in the autocorrelograms calculated from the NAP, $H1_{NAP}$, were relatively unaffected by changes in the phase-locking filter cutoff

frequency; however, the background levels at lags between the peaks were highly dependent on the cutoff frequency of the phase-locking filter. In particular, the background levels of the high-frequency band stimuli increased more than the background levels of the low-frequency band when the cutoff of the phase-locking filter was lowered. The peak-to-background ratios of the autocorrelograms of the unmodulated IRNs shown in Fig. 7 dictated the maximum dynamic range of the $R(h1_{NAP})$ of the modulated IRNs. The values of $R(h1_{NAP})$ shown in the gap regions in Fig. 6 were equal to the background levels of the respective autocorrelograms shown in Fig. 7. Therefore, the higher the background level in the autocorrelogram, the less sensitive the model is to modulations in $h1_S$.

The background levels of autocorrelograms are determined by the interaction of the half-wave rectification and lowpass-filtering processes involved in the neural transduction stage. Half-wave rectification of the basilar-membrane motion removes the negative portions of the carrier signal, shifting the mean from zero to a positive value. In the frequency domain, this positively-shifted mean manifests itself as a Fourier component at 0 Hz, referred to as a direct-current (DC) offset. In an autocorrelogram, the DC component manifests itself as an increased baseline correlation across all lags. The lowpass phase-locking filter attenuates the higher-frequency carrier-related information present in high-frequency channels more than the lower-frequency carrier-related information present in low-frequency channels. The DC component within each band, however, is unaffected by the lowpass filtering. Therefore, the peak-to-background ratio in autocorrelograms of tonal stimuli in higher frequency channels is less than the peak-to-background ratio in autocorrelograms in lower frequency channels.

The effect of the phase-locking filter on predicted thresholds can be seen in Fig. 8. The value of η was set to 7 in both low- and high-frequency bands, and the phase-locking filter cutoff frequency was varied. The predicted modulation or gap rate at which

performance reached asymptote was determined by η and therefore remained the same for stimuli in both spectral bands, irrespective of the cutoff frequency of the phase-locking filter. When the phase-locking filter was disabled, predicted thresholds from low- and high-frequency bands were almost identical at equal modulation and gap rates. As the cutoff frequency of the phase-locking filter was reduced, the thresholds predicted for the high-frequency band stimuli increased relative to the thresholds for the low-frequency band stimuli. Therefore, the phase-locking limit changed the relative sensitivity of the model to modulations in $h1_S$ between bands.

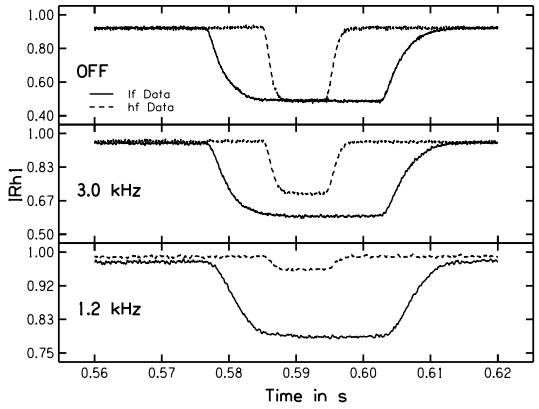

FIG. 6. Each panel shows the temporal centre of $R(h1_{NAP})$ in response to high- and low-frequency band stimuli. Solid lines are an averaged response to 1000 presentations of a 150-Hz IRN filtered into the low-frequency band, where $T_{gap}/d = 4$. Dashed lines show the same, but in response to 424.26 Hz IRNs filtered into the high-frequency band. Each panel shows data for different phase-locking filter cutoff frequencies. The integration time constant used was negligible in comparison to the gap durations so as not to affect the dynamic range of the gaps.

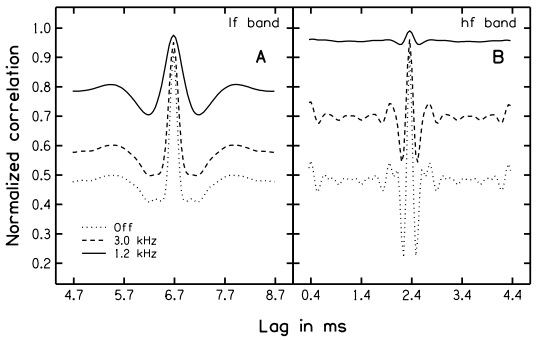

FIG. 7. Autocorrelograms of the NAP in response to unmodulated versions of the IRNs used to generate Fig. 6. These are shown because the background level relative to the peak level of the autocorrelograms of the unmodulated stimulus is indicative of the gap depth in the gap stimulus. The peak and background levels of each SACF match the peak and gap depths shown in the respective $R(h1_{NAP})$ shown in Fig. 6. The autocorrelograms presented here were calculated over the entire stimulus duration of very long IRNs (1000s). The left-hand panel shows normalized autocorrelograms in response to low-frequency band 150-Hz IRNs when the phase-locking filter was either disabled or was enabled with cutoff frequencies of 3.0 and 1.2 kHz. The right-hand panel shows the same for 424.26-Hz IRNs filtered into the high-frequency band.

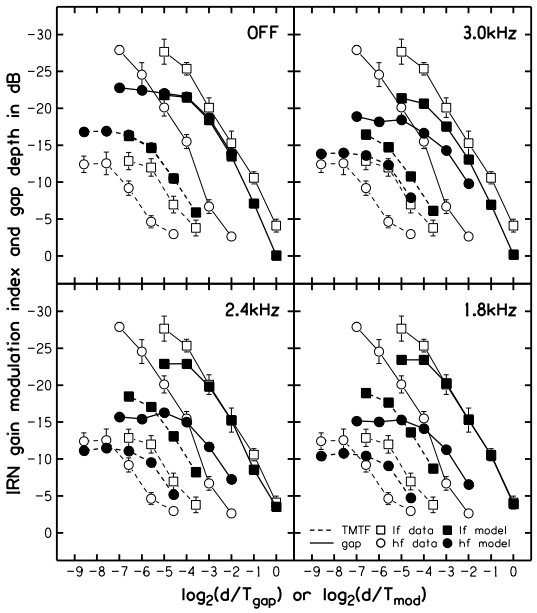

FIG. 8. Predictions from a modified version of the model used previously (Fig. 5), where η was set to 7 in both low- and high-frequency bands to isolate the effects of the phase-locking filter. Model predictions are shown superimposed upon the listener data (originally shown in Fig. 5.), where each panel shows predictions using a different phase-locking filter cutoff frequency.

The phase-locking limitation imposed a sensitivity difference between predictions of thresholds for both low- and high-frequency band stimuli. However, listeners were equally sensitive to modulations in $h1_S$, irrespective of the spectral band in which the stimuli were presented. Therefore, some form of neural compensation mechanism may be responsible for equalizing the internal representations of modulation and gap depths across frequency regions. Yost (1996) showed that the perceived pitch strength of RIN-type stimuli are monotonically related to the height of the peak occurring at the shortest non-zero lag, $H1_S$, after being subjected to an autocorrelation. Yost (1996) suggested that the function relating $H1_S$ to pitch strength is expansive, where k determines the amount of expansion.

$$E(H1_S) = \frac{10^{k*H1_S} - 1}{10^k - 1} \qquad \text{(EQN. 1.)}$$

The same kind of expansive function could be used to equalize the simulated internal ($h1_{NAP}$) representations of the gap depths between listening regions. The effectiveness of this expansive nonlinearity in equalizing gap depths is shown in Fig. 9. The upper panel shows the difference between the $h1_{NAP}$ gap depths in low- and high-frequency band stimuli as a function of k, where the parameter is the cutoff frequency of the phase-locking filter. At a supposedly more realistic cutoff frequency (~1.2 kHz), the values of k required to equalize the gap depths are large (6.4), and thus have rather severe input-output functions, as shown in the lower panel of Fig. 9.

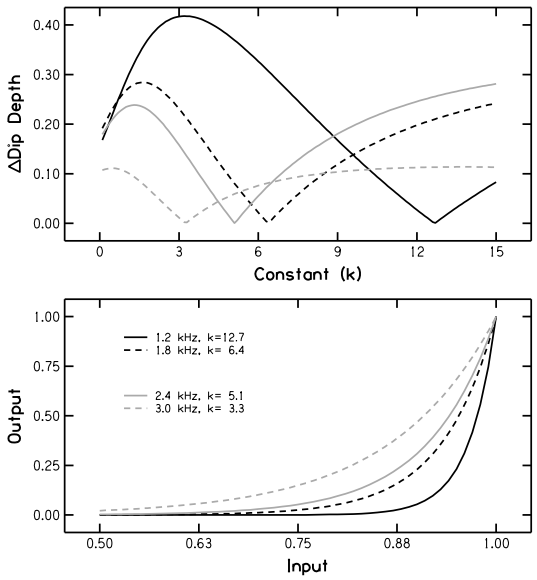

FIG. 9. The upper panel shows the mean difference between the depth of the gaps in high- and low-frequency band $R(h1_{NAP})$ (Fig. 6) after being passed through the expansive nonlinearity (Eqn. 1) as a function of the expansive constant, k. The parameter is the phase- locking filter cutoff (shown by the legend in lower panel). As the cutoff frequency of the phase-locking filter is lowered, a greater k is required to equalize the depth of the gaps. The lower panel shows the input-output relationship of the nonlinearity for each

phase-locking cutoff when the k is used that produces the minimum gap-depth difference between low- and high-band stimuli.

III. EXPERIMENT 2: MEASUREMENT OF THE TEMPORAL RESOLUTION OF PITCH PERCEPTION OVER A RANGE OF FREQUENCY REGIONS

A. Rationale

Taken together, results from Experiment 1 in the current study and results from Chapter 1 suggested that the time constants associated with neural pitch extraction scale with the interval of temporal regularity within the stimulus and also vary according to the listening region in which the stimuli are presented. However, data from the two frequency bands used so far did not provide enough information to predict η for an arbitrary frequency band. The aim of the current experiment was to measure the temporal resolution of pitch extraction over a range of frequency regions in order to gain a better picture of how η varies as a function of frequency band.

Single-value measures of temporal resolution are generally less informative than more thorough experimental paradigms, such as measurement of TMTFs, because the sensitivity of the system cannot be disambiguated from the temporal smoothing imposed by integration (Buunen and van Valkenburg, 1979). However, if one knows a priori that the system is equally sensitive across the parameter space to changes in the stimulus attribute of interest, then single-value measures are a much faster way of obtaining estimates of temporal resolution, because a single threshold is sufficient for each condition under test, as opposed to the multitude of thresholds required for a TMTF. In the current case, it is known that sensitivity is the same, because thresholds measured in the asymptotic regions of both low- and high-frequency band TMTFs were the same. Therefore, the scaling of integration times could be rapidly estimated by measuring the

shortest detectable gap in the temporal regularity of a band-limited IRN stimulus as a function of centre frequency. Increasing the centre frequency of a band-limited IRN stimulus reduces its harmonic resolvability; however, this was not an issue, as no effects of harmonic resolvability were observed in either the gap-detection data measured in the current study or in the gap-detection data presented in Chapter 1.

B. Methods

1. Stimuli

IRN stimuli were generated with a rate of 125 Hz and n=8. The stimuli were filtered into 1-kHz-wide bands using a 2^{nd}-order Butterworth filter, and thresholds were measured at centre frequencies of 0.5, 0.75, 1, 1.5, 2.5, and 3.5 kHz. Stimuli were 1024 ms in duration, presented at a level of 65m dB SPL, and were gated on and off with 16-ms cosine-squared ramps. Stimuli were presented in a continuous noise to mask audible distortion products below the stimulus passband. This noise was lowpass-filtered at 0.5 octaves below the lower cutoff frequency of the stimulus, prior to which the noise was filtered in the spectral domain so as to produce a roughly constant excitation level of 30-dB SPL per equivalent rectangular bandwidth.

2. Procedure

Each trial consisted of two observation intervals, which were separated by 500-ms gaps. One interval contained an IRN with a gap in $h1_S$, while the other interval contained an IRN with no gap. The listeners' task was to detect the interval containing the gap. Intervals were presented in a random order within each trial. In the target intervals, the gap was positioned symmetrically around the temporal centre of the stimulus. An adaptive staircase technique was used to measure thresholds where the adaptive parameter was the

duration of the gap. At the beginning of each threshold run, the gap duration was much longer than the anticipated detection threshold. The gap duration was decreased after three consecutive correct responses and increased after each incorrect response. The ratio for the increments and decrements in gap duration was 2 for the first reversal in level, 1.5 for the second reversal, and 1.25 for the remainder of the 10 reversals that made up each threshold run. The last 8 reversals of the gap duration were geometrically averaged to obtain a threshold estimate for each run. Three threshold runs were conducted for each participant per stimulus condition using a counter-balanced design to eliminate training effects.

3. Listeners

A total of 5 listeners (3 male and 2 female, aged between 21 and 33 years) participated in Experiment 2. Listeners met the same criteria as outlined in Experiment 1.

C. Results

The results of this experiment are shown in Fig. 10. When plotted on log-log axes, the threshold pattern resembles an inverted lowpass filter function. Mean thresholds increased with increasing listening region from around 4 ms when the stimulus contained frequencies between 0 and 1 kHz, to just over 50 ms when the stimulus contained frequencies between 3 and 4 kHz. Integration time constants cannot be directly derived from the data presented in Fig. 10. However, the gap-duration thresholds can be considered proportional to the neural integration time constants in those listening regions, as no frequency-region-dependent differences in sensitivity were observed in the TMTF data.

The statistical significance of these results was tested using a repeated-measures ANOVA performed on the factor centre frequency, from which a significant main effect

of centre frequency was observed [F(5,20)=46.346, p<0.001]. Pairwise comparisons between thresholds at consecutive centre frequencies showed that while differences between thresholds for stimuli presented in bands centred at 0.50 and 0.75 and also 0.75 and 1.00 kHz were not significantly different, thresholds for stimuli presented in bands centred at 0.50 and 1.00 were significantly different [F(5,20)=46.346, p=0.025]. Thresholds for stimuli presented in bands centred at 1.00 kHz and above were all significantly different (between 1.00 and 1.50 kHz [F(5,20)=46.346, p<0.001], between 1.50 and 2.50 [F(5,20)=46.346, p=0.014], and between 2.50 and 3.50 [F(5,20)=46.346, p=0.027]).

To characterize the roll-off of the significantly different thresholds at consecutive centre frequencies between 1.00 and 3.50 kHz, a log-log regression was performed on the mean data. The function describing the roll-off of the lowpass filter relative the centre frequency of the band in which the stimulus was presented, cf, was best approximated by $7.60 \cdot cf^{-1.56}$, the rejection rate of which can be quantified as 9.4 dB per octave or 31.2 dB per decade.

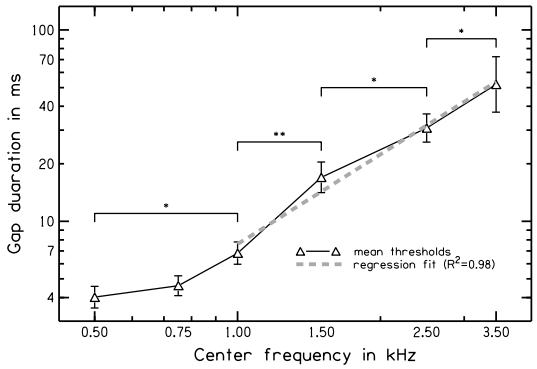

FIG. 10. Thresholds for the shortest detectable gap in temporal regularity in an IRNO (d=8, n=8), averaged across the 5 listeners and plotted as a function of centre frequency. A regression line is also plotted between centre frequencies between 1.00 and 3.50 kHz.

Error bars represent inter-listener standard error. Square brackets grouping data points show where the mean significance is different at the 0.05 level (*) and the 0.01 level (**).

IV. DISCUSSION

As in Chapter 1, the harmonic resolvability of the high-frequency band stimuli used in the current study appeared to have no effect on the temporal resolution of pitch extraction. Again, like in Chapter 1, this suggests that pitch is extracted by a temporal mechanism, even though the TFS information would be expected to be degraded in the high-frequency band used. Alternatively, the lack of resolvability effect suggests that spectral and temporal pitch- extraction mechanisms share common or functionally similar integration processes.

Results from the current study suggest that the pitch-extraction mechanism uses far longer integration windows in higher frequency regions compared to lower frequency regions. This means that the system loses temporal acuity and rapid changes in temporal regularity become less detectable in higher frequency bands. The value of η derived from the high-frequency band TMTFs measured in the current study was 17. However, the pitch-discrimination data of White and Plack (2003) measured in a similar frequency region (2.75 to 3.75 kHz) suggested that pitch information was only integrated over ~10 stimulus cycles. This seeming paradox is investigated in detail in Chapter 4.

The asymptotic thresholds in TMTF measurements reveal the sensitivity of the system to the modulations in $h1_S$. The asymptotic thresholds of the TMTF measurements made in the current study (which were not confounded by the bandpass characteristic) were approximately -14dB. Therefore, the sensitivity of the auditory system to modulations in $h1_S$ was not affected by the repetition rate of the stimulus. Moreover, the asymptotic thresholds in the TMTFs measured in Chapters 1 and 2 were very similar,

suggesting that the sensitivity of the system to modulations in $h1_S$ was equal in both spectral bands.

The autocorrelation-based pitch-extraction model presented in the current study was successfully able to simulate the observed loss of temporal acuity in the high-frequency data measured in the current chapter relative to the low-frequency data measured in Chapter 1. This was achieved by using larger integration time constants in the higher frequency channels of the model. However, the autocorrelation-based model predicted that the system would be less sensitive to modulations in $h1_S$ in the high-frequency band. This was attributed to the interaction of the half-wave rectification and lowpass filtering used to simulate neural transduction. To compensate for this loss of sensitivity, a fix was suggested, whereby the simulated internal representation of pitch strength was passed through an expansive function before the decision mechanism. However, in order to minimize sensitivity differences between spectral bands, this expansive function needed to be rather severe ($k=6.4$) in comparison to that known to relate $H1_S$ to the perceived pitch strength associated with IRN stimuli ($k\sim=1$) (Wiegrebe et al., 1998). The implementation of the expansive function in a neural model of pitch strength is investigated in more detail in Chapter 3.

The final experiment in the current chapter measured the temporal resolution of pitch extraction over a wide range of frequency regions to get a better idea of how η varies with frequency band. The mean threshold pattern measured in Experiment 2 resembled an inverted lowpass filter with a cutoff in the region of 1.00 kHz with a rejection rate somewhere between that of a 1^{st}- and 2^{nd}-order filter. Phase-locking accuracy in humans has been inferred from behavioural studies measuring the upper frequency limit for interaural phase-difference (IPD) detection (Garner and Wertheimer, 1951, Ross et al., 2007, Schiano and Trahiotis, 1985, Zwislocki and Feldman, 1956), which has generally

been shown to be between 1.1 and 1.3 kHz. The phase-locking limitations of the neural transduction process are commonly modelled as a 2^{nd}-order lowpass filter with a cutoff around 1.2 kHz. Therefore, the gap-detection data may reflect the system using longer integration times to compensate for the progressive loss of high-fidelity temporal fine-structure information towards higher frequency regions. This novel behavioural task may be an effective method for monaural quantification of the breakdown of phase locking in humans. Furthermore, the data suggests that while the degraded TFS available in high-frequency channels may be of little use for IPD detection, it can still be integrated and utilized for pitch extraction.

Chapter 3

The temporal resolution of pitch perception III: Effects of pitch strength

I. INTRODUCTION

In Chapters 1 and 2, the temporal resolution of the auditory system was measured in response to changes in the instantaneous temporal regularity within the stimulus. For this, a novel stimulus based on iterated ripple noise (IRN) was used that allowed the instantaneous temporal regularity, $h1_S$, to be changed over time. This new stimulus enabled measurement of pitch-domain gap-detection thresholds and temporal modulation transfer functions (TMTFs). TMTFs are a particularly useful measure of temporal acuity, as the sensitivity of the system to the modulations can be disentangled from the temporal smoothing effects imposed by peripheral and neural integration processes (Viemeister, 1979).

Results from Chapters 1 and 2 showed that the system was equally sensitive to modulations in $h1_S$, irrespective of the rate of the IRN stimulus. This finding was unexpected, as temporal models of pitch generally apply a function that weights the autocorrelogram less and less towards longer lags, and this would imply that the system should be less sensitive to modulations in $h1_S$ at lower IRN rates. However, this weighting function was based on the results of pitch-value discrimination tasks measuring resolution towards the lower limit of pitch. Furthermore, results from a magnitude-estimation task (Fasti, 1988) showed that the subjective pitch strength of an IRN stimulus was also dependent on the rate of the stimulus. Again, this would imply that the system should be less sensitive to modulations in $h1_S$ at lower IRN rates. However, none of these tasks measured the sensitivity of the system to changes in pitch strength as was measured in Chapters 1 and 2.

Comparison of results from Chapters 1 and 2, in which stimuli were presented in low- and high-frequency regions, revealed that the system was equally sensitive to modulations in temporal regularity, irrespective of the listening region in which the stimuli

were presented. Studies using IRN stimuli have shown that spectral peaks numbered 3 to 5 dominate the percept of both pitch value and pitch strength associated with the stimuli (Yost, 1982, Yost and Hill, 1978). The dominance measurements would suggest that lower harmonics (and thus IRNs presented in lower frequency regions) elicit stronger pitch. This would imply that the system should be less sensitive to modulations in $h1_S$ in stimuli presented in higher frequency regions, but again, the dominance measurements are based on subjective judgements of pitch strength.

Chapters 1 and 2 showed that the pitch-integration time constants depend on the pitch value of the stimuli, and pitch value is known to have an effect on the subjective pitch strength of the stimuli. Comparison of results from Chapters 1 and 2 showed that the integration-window time constants depend on the frequency region in which the stimuli are presented and frequency region is also known to have an effect of subjective pitch strength associated with the stimuli. The aim of the current study was to see whether the time constants of pitch extraction depend on the subjective pitch strength of the stimulus when the pitch strength is varied by changing the number of iterations, n, used in the IRN circuit, rather than changing the frequency range in which the IRN stimuli are presented.

Results from this part of the study showed that thresholds are higher for lower n, but function shapes suggest that this difference is mainly a difference in sensitivity: lower n translates to an overall lower $H1_S$, and so, listeners would be expected to perform worse under these conditions. However, when thresholds were plotted in units of $h1_S$, there was still a sensitivity difference between thresholds for stimuli with different n.

In an earlier study, Yost et al. (1996) showed that the pitch strength that listeners associate with an IRN stimulus is monotonically related to the height of the first-order peak in the autocorrelogram of the stimulus, $H1_S$. In a subsequent study, Yost (1996) used a magnitude-estimation method to relate the perceived pitch strength of IRNs to their $H1_S$

and suggested that pitch strength is related to an expanded representation of $H1_S$.Wiegrebe et al. (1998) compared the pitch strength associated with a rippled noise (RN) to a repeated-period noise (RPN) stimulus with an $h1_S$ that was square-wave modulated between 0 and 1 at rates above the modulation-detection threshold. At these high modulation rates, the RPN was perceived to have static pitch strength and a tonal quality similar to that of the RN stimulus. However, the modulated RPN stimulus elicited greater pitch strength than the unmodulated RN stimulus, even though both stimuli had an overall $H1_S$ of 0.5. This result was explained by assuming that $h1_s$ is integrated after being subjected to an expansive nonlinearity. Therefore, the average expanded $h1_s$ of the modulated stimulus was greater than that of the unmodulated stimulus. The results of the current study could be modelled using this expansive function and time constants that were independent of the subjective pitch strength of the stimulus. The second part of the current study considers implications of cochlear compression on how expansion should be modelled in a neural model of pitch strength.

II. METHODS

A. Stimuli

TMTF and gap-detection thresholds were measured using the modified IRN circuit presented in Chapter 1. The main parameter in this study was the average temporal regularity of the stimuli, $H1_S$, which was adjusted by changing the number of iterations, n, in the IRN circuit. Thresholds were measured for IRNs with n = 1, 2, 4, and 8. This allowed for quantification of whether the time constants of pitch extraction are dependent on the overall $H1_S$ of the stimulus. As in Chapters 1 and 2, harmonic resolvability of the stimuli was an experimental parameter. As in Chapter 1, thresholds were measured for IRNs filtered into a 2.2-kHz bandwidth with a centre frequency of 1.88 kHz using IRN

rates of 53.05 Hz (unresolved) and 106.07 Hz (resolved). Stimuli were presented at a level of 65 dB sound pressure level (SPL) and were gated on and off with 5-ms cosine-squared ramps. Stimuli were presented in a continuous noise to mask audible distortion products below the stimulus passband, using the same methods and equipment as in Chapter 1.

B. Procedure

Gap-depth thresholds were measured for gap durations (T_{gap}) equal to multiples of 1, 2, 4, 8, 16, 32, and 64 times each IRN delay, d. Modulation-detection thresholds were measured for modulation periods (T_{mod}) equal to multiples of 6, 12, 24, 48, and 96 times each IRN rate. The stimulus durations were 1.2068 seconds in the gap experiment and 1.8102 seconds in the modulation experiment as described in Chapter 1. This allowed for at least one complete modulation cycle of the slowest modulation rate used. Gap- and modulation-detection thresholds were measured using the adaptive procedure described in Chapter 1; again, the adaptive parameter was the gap depth or modulation index defined in terms of the IRN circuit gain, g.

Informal listening revealed that pitch-strength fluctuations in some of the shorter T_{gap} and T_{mod} conditions were not detectable when the number of iterations of the IRN was less than 8, even when the depth of the modulation or gap was maximum (0 dB). Therefore, if a listener was unable to detect a particular gap or modulation rate for a certain n on more than 2 consecutive occasions, that condition (and any shorter T_{gap} or T_{mod}) was considered un-measurable for that individual and not tested again.

C. Listeners

A group of 8 listeners, who were different to those who participated in the companion studies, participated in the current experiments. One subset of 4 listeners (all

female, aged between 23 and 26 years) participated in the gap-detection experiment, and the other subset of 4 listeners (2 male and 2 female, aged between 24 and 37 years), one of whom was the author, participated in the modulation-detection experiment. Participants were paid for their services at an hourly rate and met the criteria outlined in Chapter 1.

III. RESULTS AND INTERIM DISCUSSION

A. Thresholds represented in terms of the adaptive parameter, g

Average detection thresholds are shown in Fig. 1, where thresholds are plotted in units of the adaptive parameter, gap depth, or modulation index defined in terms of g. As in Chapters 1 and 2, data are plotted with axes reversed so that threshold patterns resemble lowpass filter functions. The statistical significance of the observations was tested by performing linear mixed-models analyses on both modulation- and gap-detection data. For the gap-detection task, the analysis was performed on factors T_{gap}/d, IRN rate, and n. For the modulation-detection task, the analysis was performed on factors T_{mod}/d, IRN rate, and n. The dependent variable was the mean average threshold for each participant in each condition. Thresholds were significantly higher at shorter gap durations, as shown by the significant main effect of T_{gap}/d [$F(6,93.046)=142.240$, $p<0.001$]. Similarly, thresholds were significantly higher at greater modulation rates, as shown by the significant main effect of T_{mod}/d [$F(4,65.043)=54.996$, $p<0.001$]. Only the best performing listeners were able to obtain thresholds for the shortest gaps and highest modulation rates measured. Asterisks adjacent to some data points in each panel denote the number of listeners who were unable to obtain a threshold in those conditions. The modulation detection task where n=1 and the IRN rate was 53.03 Hz was so difficult that no listeners were able to obtain a threshold, even at the slowest modulation rates attempted. There was a significant main effect of n for both the gap- [$F(3,93.025)=249.027$, $p<0.001$] and modulation-

detection experiments [$F(3,65.014)=95.711$, $p<0.001$], in that listeners were able to obtain lower thresholds for IRNs with greater n at equal modulation rates and gap durations. However, the threshold patterns look similar for IRNs with different n, suggesting that the observed differences are mainly due to differences in sensitivity rather than differences in integration time. The pitch strength associated with an IRN is monotonically related to its $H1_S$, and IRNs with lower n have an overall lower $H1_S$. This means that the dynamic range of the modulations and gaps in terms of $h1_S$ is lower for stimuli with lower n; therefore, listeners would be expected to perform worse. In the next analysis, thresholds are plotted in terms of $h1_S$.

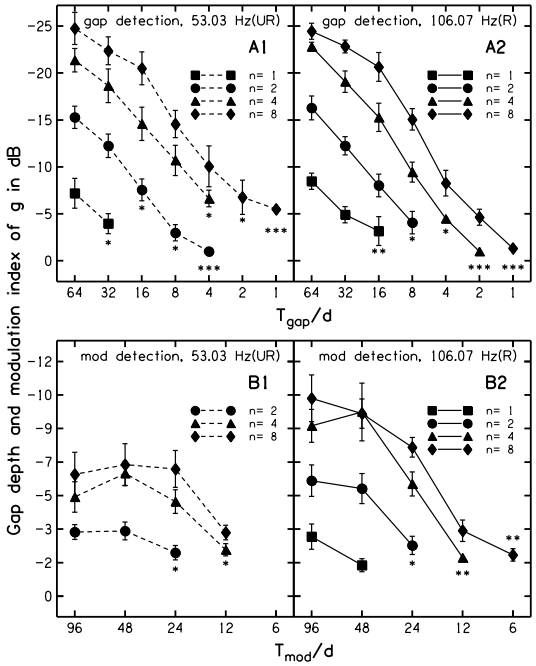

FIG. 1. Experimental results where thresholds are plotted in terms of the adaptive parameter, g. Thresholds are averaged across 4 listeners, where the error bars represent inter-listener standard error. Upper panels show gap-detection results, where the ordinate is the gap duration (T_{gap}) normalized by the IRN delay, d. Lower panels show modulation-detection results, where the abscissa is the modulation period (T_{mod}) normalized by the IRN delay, d. The axes are reversed in each case so that the TMTF results resemble low-pass filter functions. Left-hand panels show thresholds for unresolved (UR) IRNs, and right-hand panels show thresholds for resolved (R) IRNs. The parameter in each panel is n. Asterisks adjacent to some data points represent the number of listeners who were unable to obtain a threshold in those conditions.

B. Thresholds represented in terms of $h1_S$

The instantaneous $h1_S$ associated with the stimulus at a given point in time, t, is related to g and n by Eqn. 1.

$$h1_S(t) = \frac{n}{n+1} \cdot g(t) \qquad \text{(EQN. 1.)}$$

Therefore, a gap-depth threshold in terms of g, g_D, can be converted to an $h1_S$ gap-depth threshold, $h1_{SD}$, by calculating the difference in between the $h1_S$ outside and within the region of the gap, as shown by Eqn. 2.

$$h1_{SD} = \frac{n}{n+1} \cdot g_D \qquad \text{(EQN. 2.)}$$

Similarly, a modulation index threshold in terms of g, g_m, can be converted to an $h1_S$ modulation index threshold, $h1_{Sm}$, by calculating the difference in between the $h1_S$ at the peaks and minima of the modulations, as shown by Eqn. 3.

$$h1_{Sm} = \frac{n}{n+1} \cdot g_m \qquad \text{(EQN. 3.)}$$

Fig. 2. shows the listener thresholds converted into units of $h1_S$. The statistical significance of the observations was tested once more using the analysis described above, but where the dependent variable was the mean threshold for each participant in each condition in units of $h1_S$. While the functions for different n look more compressed relative to each other compared to when thresholds were plotted in terms of g (Fig. 1.), there was still a highly significant main effect of n in both the gap-detection [$F(3,93.025)=134.176$, $p<0.001$] and modulation-detection [$F(3,65.014)=13.870$, $p<0.001$] tasks.

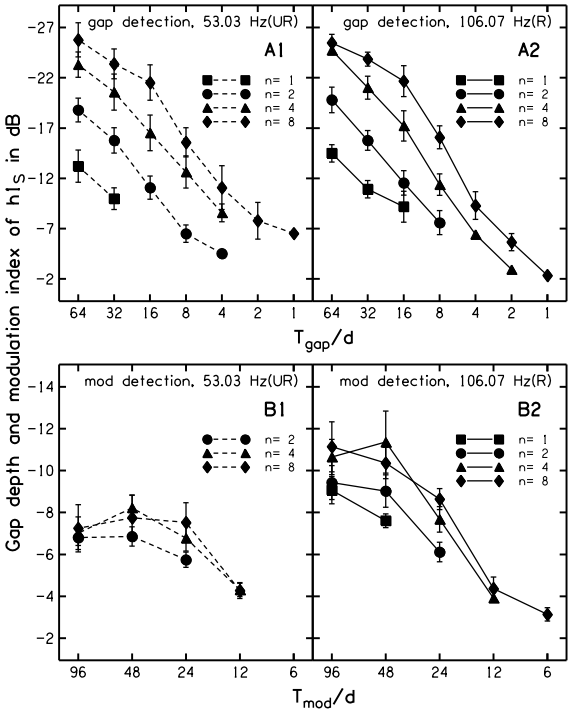

FIG. 2. Data from Fig. 1. plotted in $h1_S$ units. Again, the upper panels show the gap-detection thresholds. The lower panels show the modulation-detection thresholds.

C. Thresholds represented in terms of E($h1_S$)

When thresholds were converted into units of $h1_S$, the sensitivity difference between stimuli with different n was still present. This could suggest that the expansive relationship between $H1_S$ and pitch strength (Yost 1996) also applies to listeners' sensitivity to modulations in $h1_S$ over time. To see if such an expansive function is able to eliminate the observed effect of n, thresholds were converted into expanded $h1_S$ units, E($h1_S$). The expansive function used, E, is defined by Eqn. 4, where the constant, k, determines the expansiveness.

$$E(h1_S) = \frac{10^{k*h1_S} - 1}{10^k - 1} \qquad \text{(EQN. 4.)}$$

Therefore, a gap-depth threshold in terms of g, g_D, (as plotted in Fig. 1.A) can be converted to a gap-depth threshold in terms of E($h1_S$) units, E($h1_{SD}$), by calculating the

difference between the $E(h1_S)$ outside and within the region of the gap, as shown by Eqn. 5.

$$E(h1_{SD}) = E\left(\frac{n}{n+1}\right) - E\left(\frac{n}{n+1}\cdot(1-g_D)\right) \qquad \text{(EQN. 5.)}$$

Similarly, a modulation index threshold in terms of g, g_m, (as plotted in Fig. 1.B) can be converted to a modulation index threshold in terms of $E(h1_S)$, $E(h1_{Sm})$, by calculating the difference in between the Eh1 at the peaks and the Eh1 at the minima of the modulations, as shown by Eqn. 6.

$$E(h1_{Sm}) = E\left(\frac{n}{n+1}\cdot\frac{1+g_m}{2}\right) - E\left(\frac{n}{n+1}\cdot\frac{1-g_m}{2}\right) \qquad \text{(EQN. 6.)}$$

The effectiveness of the expansive function at minimizing sensitivity differences between thresholds for stimuli with different n was determined by calculating the sum of the variance between thresholds at each modulation or gap rate as a function of k. For this, individual mean thresholds for each listener were converted into $E(h1_S)$ units using values of k spaced linearly between 0.2 and 2.0 in steps of 0.1. For each value of k, $E(h1_S)$ thresholds were averaged across listeners to give a mean $E(h1_S)$ threshold, which was then converted into dB. The standard deviation of the mean thresholds in response to stimuli with different n at each T_{gap}/d and each T_{mod}/d was summed to give an error score for each value of k. Results of this analysis are shown in Fig. 3. The overall error score when averaged across both IRN rate and task was lowest when k=1.2.

Examined separately, the error scores for the gap-detection data (circles) had well-defined minima occurring at k=1.2, whereas the minima for the TMTF data (squares) occurred at lower values of k and were not so well-defined. Compared to the gap-detection experiment, listeners reported that the modulation-detection experiment was more difficult, and the inter-listener error in the modulation-detection thresholds was higher

than that in the gap-detection thresholds, on average. Therefore, the modulation-detection data was less useful for defining k.

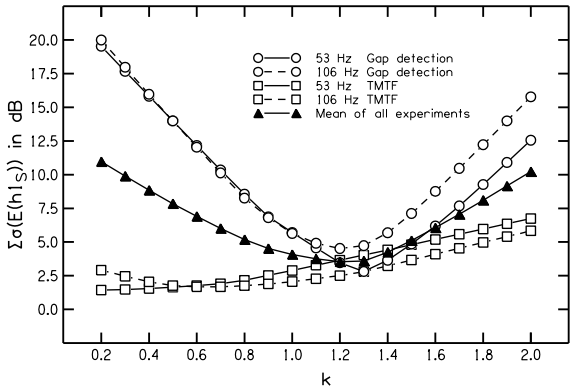

FIG. 3. The mean of the standard deviation of thresholds for each T_{mod}/d and each T_{gap}/d after being processed by the expansive nonlinearity as a function of the expansive constant, k, used in Eqn. 2. This is shown separately for each IRN rate in both gap- and modulation-detection tasks. The mean of standard deviation across all experiments is also displayed.

Using the best value of k=1.2, threshold patterns from both gap- and modulation-detection tasks in units of $E(h1_S)$ are shown in Fig. 4. Plotting gap-detection thresholds in terms of $E(h1_S)$ accounts for almost all of the sensitivity differences between thresholds in response to stimuli with different n. The statistical significance of observations was tested once more by performing linear mixed-models analyses. The dependent variable was the mean threshold for each participant in each condition in units of $E(h1_S)$. For the gap-detection data, only the main effect of T_{gap}/d remained significant. Therefore, the expansive function was able to account for the main effect of n observed when thresholds were represented in units of g and $h1_S$. For the modulation-detection data, the main effect of n was still significant at the 0.05 level [$F(3,65.014)=4.020$, $p=0.011$]. There was no

significant interaction between IRN rate and n, [F(2,64.997)=0.865, p=0.426] indicating that the effect of n was similar for both rates.

Careful inspection of the TMTF data in Fig. 4 suggested that the expansive function could account for sensitivity differences between thresholds in the roll-off regions, but not the asymptotic regions of the TMTFs. Pairwise comparisons between thresholds at each T_{mod}/d revealed that thresholds were insignificantly at $T_{mod}/d < 48$. At $T_{mod}/d = 48$ thresholds were significantly different [F(3,64.994)=5.658, p=0.002], and at $T_{mod}/d = 96$ thresholds were significantly different [F(3,64.994)=11.641, p<0.001]. In fact, the expansion overcompensated for the sensitivity differences in the asymptotic region, probably because performance was limited more by the stimulus duration than by n in this region. However, the gap-detection data strongly suggests that the internal decision mechanism is based on an expanded representation of $h1_S$.

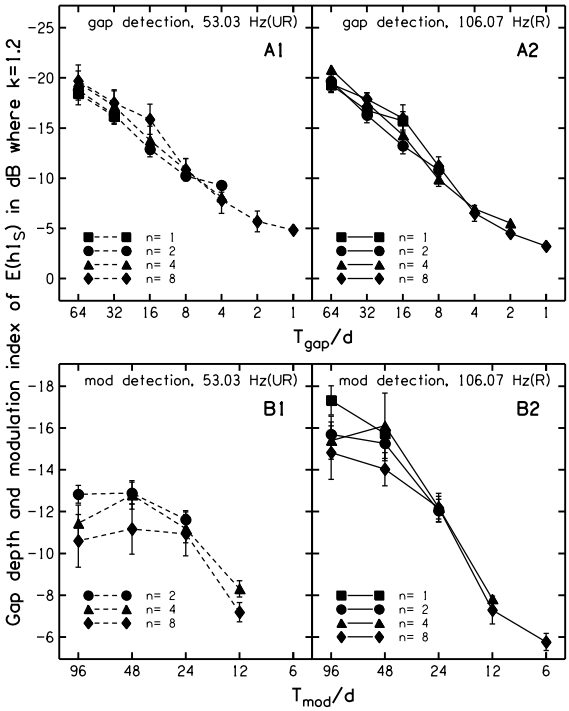

FIG. 4. Thresholds are converted into $E(h1_S)$ units, where k=1.2. The upper panels (A) show the gap-detection thresholds. The lower panels (B) show the modulation-detection thresholds.

IV. TOWARDS AN IMPROVED TEMPORAL MODEL OF PITCH STRENGTH

Data from the current study suggests that our sensitivity to modulations in temporal regularity is based on the expanded instantaneous temporal regularity of the signal, $E(h1_S)$, where k=1.2. Similarly, Yost (1996) has shown that the pitch strength that listeners associate with IRNs is proportional to $E(H1_S)$. However, the model presented in Chapter 2 showed that simulated nonlinear processes in the auditory periphery increased the baseline correlation in the autocorrelogram of the NAP and so the input-output (I/O) function relating $h1_S$ to $h1_{NAP}$ was compressive.

The autocorrelogram of the NAP in response to a stimulus can be used to quantify the maximum dynamic range of fluctuations in $h1_{NAP}$ because, theoretically, the $h1_{NAP}$ of a modulated stimulus is bounded by the $H1_{NAP}$ and the background level of the

autocorrelogram of an unmodulated stimulus (for a detailed discussion, refer to Chapter 2). To illustrate the output of this model, autocorrelograms of the NAP were generated for 106.07-Hz IRNs. The autocorrelograms are shown in Fig. 5, where the parameter is the phase-locking cutoff frequency. As described in Chapter 2, the background levels of the autocorrelograms increase relative to the peak ($H1_{NAP}$) when the cutoff frequency of the phase-locking filter is lowered. To quantify the compressive relationship between $h1_S$ and $h1_{NAP}$, I/O functions were generated where the parameter was the phase-locking cutoff frequency. For this, the $H1_S$ of an IRN was adjusted in linear increments, from 0 to maximum by incrementing g in linear steps. The $H1_{NAP}$ was then recorded for each value of g. Comparison of the upper panels of Fig. 6 shows that the I/O functions (right-hand panel) were bounded by the peak and background levels of their corresponding autocorrelograms (left-hand panel). As the phase-locking cutoff frequency of the phase-locking filter was lowered, $H1_{NAP}$ was compressed more and more relative to $H1_S$. Also plotted in the right-hand panel is the I/O function of $E(H1_S)$, using the value of k=1.2 derived from the data measured in the current study. The I/O functions generated from the NAPs were compressive, whereas $E(H1_S)$ (on which the data measured in the current study is thought to be based) is expansive.

The I/O functions generated from the NAPs were then subjected to the expansive process (Eqn. 4), and the RMS deviation between I/O functions relating $H1_S$ to $E(H1_S)$ and $H1_S$ to $E(H1_{NAP})$ were plotted as a function of k in the lower panel of Fig. 5. As the phase-locking cutoff frequency was lowered, an increasingly higher k was required in the expansive function to map $E(H1_{NAP})$ to $E(H1_S)$. Data from Experiment II in Chapter 2 suggested that the phase-locking cutoff frequency should be modelled using a value in the region of 0.8 to 1.2 kHz. Use of a value in this range would require an exceptionally

severe (k > 5.5) internal expansive function in order to model the data presented in the current study.

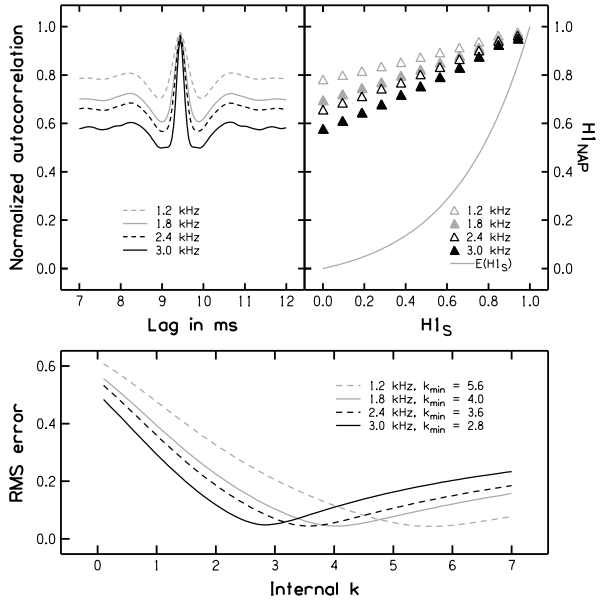

Fig. 5. Parametric effects of phase-locking filter cutoff frequency on the dynamic range of $h1_{NAP}$. The upper-left panel shows the long-term autocorrelation functions of the neural activity patterns in response to an IRN with a rate of 106.07 Hz. The plot is centred on the first peak of the autocorrelogram, and the parameter is the cutoff frequency of the phase-locking filter. The upper right-hand panel shows the I/O functions relating $H1_S$ to $H1_{NAP}$ as g was varied between 0 and 1. Once again, the parameter is the cutoff frequency of the phase-locking filter. For reference, the I/O function relating $H1_S$ to $E(H1_S)$ where k= 1.2 is shown as a solid line in the same panel. The bottom panel shows the RMS difference between the I/O functions after processing with an expansive nonlinearity, and $E(H1_S)$, plotted as a function of k. Again, the parameter is the cutoff frequency of the phase-locking filter.

Models of the auditory periphery generally involve at least a simple, instantaneous-compression process to coarsely simulate cochlear compression. Compression is a nonlinear process that is also likely to affect the relationship between $H1_S$ and $H1_{NAP}$. The aim of the current analysis is to consider the implications of cochlear compression on how the expansive process should be modelled in a neural model of pitch strength. This is done by defining the relationship between $H1_S$ and $H1_{NAP}$ when using different compression schemes. The compression schemes tested included a linear gammatone filter bank with logarithmic ($\log_{10}$), power-law ($x^{1/2}$), and $x^{1/8}$ compression (where x represents the signal within each channel). A dynamically compressive cochlear model, the pole-zero filter cascade (PZFC) (Walters, 2010), was also used.

When using a linear filter bank such as the gammatone, compression is often applied as a simple instantaneous power law or logarithmic compression scheme, as implemented in Chapters 1 and 2. More recent functional cochlear models, such as the PZFC, provide compression in a dynamic and thus more realistic manner. A block diagram of the PZFC is shown in Fig. 6. The PZFC applies a variable gain to the signal within each channel that results in a compressed output relative to its input. The adaptive gain control (AGC) is temporally dynamic, reflecting the time course of efferent feedback processes that regulate the gain. The AGC is also mediated by activity in neighbouring channels to account for two-tone suppression data (Sachs and Kiang, 1968). The parameters that govern the behaviour of the AGC were fitted to psychoacoustical notched-noise masking data (Glasberg and Moore, 2000).

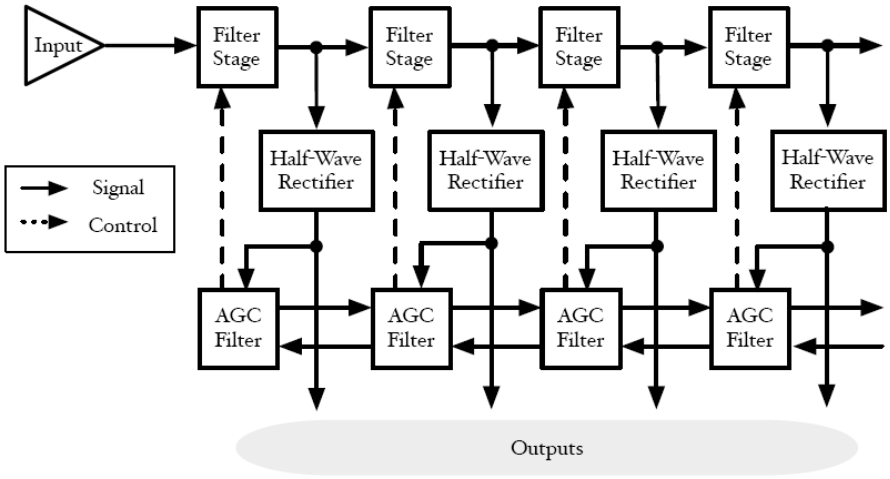

Fig. 6. Copy of Fig. 2.2 from Walters (2010), shown with permission. Flow diagram illustrating the audio-signal and control-signal paths within the PZFC cochlear model. The input signal cascades through each filter stage, where the gain associated with each filter is mediated by an adaptive gain control mechanism. The adaptive gain is mediated by both the signal within the associated channel and signals from neighbouring channels.

Before the effects of a specific compression scheme on temporal regularity can be interpreted, one must first define how different compression schemes compress the level of the input signals. For this, signal level I/O functions were generated by measuring the RMS output level of the model's peripheral channel centred closest to 1 kHz in response to a sinusoidal input signal of corresponding frequency over a range of input levels. The resulting level I/O functions from each compression scheme are plotted in the upper panel of Fig.7. Plotting the derivative of the I/O functions with respect to input level (Fig. 7, lower panel) provided the compression ratio in terms of dB output per dB input. Power-law compression gives a constant compression ratio of 1/2-dB output per 1-dB input, irrespective of input level. Similarly, $x^{1/8}$ compression gives 1/8-dB output per 1-dB input, irrespective of input level. In contrast, the compression ratios of both logarithmic and

PZFC compression are level-dependent. The RMS level of the stimuli used in the current study was 0.1. At this input level, the logarithmic compression has a very similar compression ratio to the $x^{1/8}$ compression and the PZFC has a very similar compression ratio to the power-law scheme. This allowed testing of whether compression schemes with similar compression ratios have similar effects on the compressive relationship between $H1_S$ and $H1_{NAP}$.

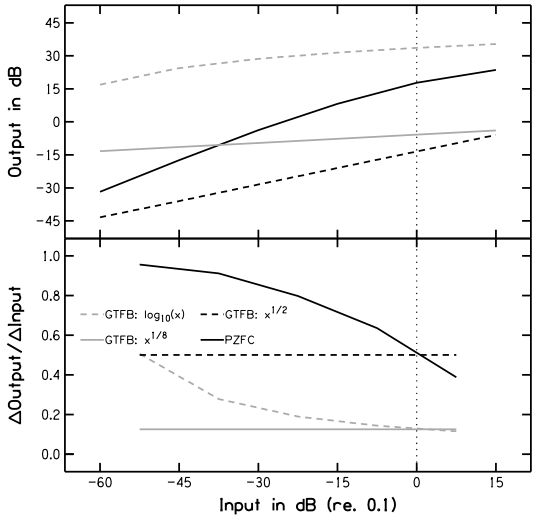

FIG.7. Analysis of the compression characteristics used in the auditory model. The upper panel is the RMS amplitude I/O function of the 1007-Hz channel of the auditory model in response to a sinusoidal stimulus of the same frequency. The parameter is the compression scheme. The vertical dashed line represents the equivalent RMS level at which the stimuli were input into the model in subsequent simulations (RMS input level = 0.1). The lower panel shows the first differential of the simulated data in the upper panel with respect to input level, giving the dB output per dB input for each input level. At 0-dB relative input, logarithmic compression is equivalent to $x^{1/8}$, while PZFC compression is equivalent to square-law compression.

To test the effects of the compression scheme on $H1_{NAP}$, Fig. 8 was generated using the same methods used to generate Fig. 5, but where compression scheme was the parameter and the phase-locking cutoff frequency was set to a lenient value of 3.0 kHz in order to emphasize the effects of compression over phase-locking limitations. Focusing on the upper left-hand panel, it is evident that the autocorrelograms generated using logarithmic and $x^{1/8}$ compression schemes are very similar and therefore in agreement with the hypothesis that compression schemes with similar compression ratios may have similar effects on the compressive relationship between $H1_S$ and $H1_{NAP}$. However, autocorrelograms generated using power-law and PZFC compression schemes were different to each other, despite having similar compression ratios in response to a sinusoid. In particular, the background level of the PZFC autocorrelogram was considerably lower than the background level of the power-law autocorrelogram, and thus the I/O function relating $H1_S$ to $H1_{NAP}$ was much less compressive. This was also reflected in the value of k required to map the I/O function relating $H1_S$ and $H1_{NAP}$ to the I/O function relating $H1_S$ and $E(H1_S)$. The value of k required when using the PZFC was substantially less than the value of k required when using instantaneous-compression schemes, as shown by the bottom panel of Fig. 8.

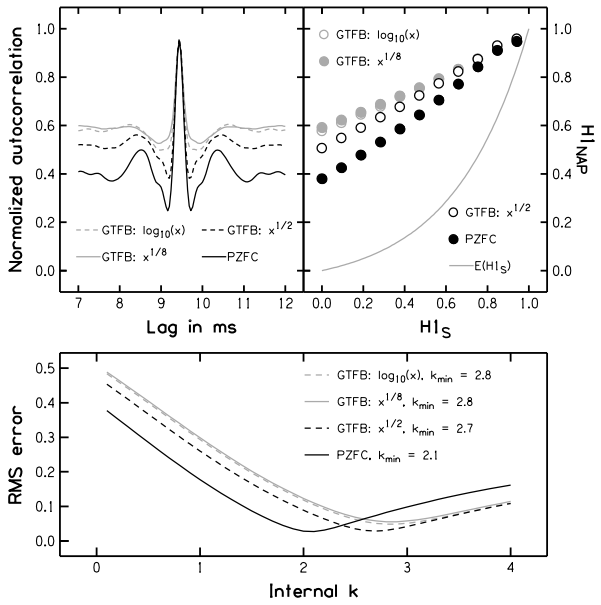

FIG. 8. Parametric effects of peripheral compression scheme on $H1_{NAP}$. The upper-left panel shows the long-term autocorrelation functions of the neural activity patterns in response to an IRN with a rate of 106.07 Hz. The plot is centred on the first peak of the ACF, and the parameter is the compression scheme used. The upper right-hand panel shows the I/O functions relating $H1_S$ to $H1_{NAP}$ as g was varied between 0 and 1. For reference, the I/O function relating $H1_S$ to $E(H1_S)$ where k= 1.2 is shown as a solid line in the same panel. The bottom panel shows the RMS deviation between the I/O functions describing $E(H1_S)$ and $E(H1_{NAP})$ over a range of k used in $E(H1_{NAP})$.

The data shown in Fig. 5 demonstrated that the compressive relationship between $H1_S$ and $H1_{NAP}$ became increasingly more compressive as the phase-locking filter cutoff frequency was reduced. To investigate potential interactions between the intensity compression scheme and the cutoff of the phase-locking filter, k was derived to map $E(H1_{NAP})$ to $E(H1_S)$ for each combination of compression scheme used in Fig. 8 with each cutoff frequency used in Fig. 5. The results of this are shown in Fig. 9. The relative values of k required when using instantaneous-compression schemes increased proportionally to

one another as the phase-locking cutoff frequency decreased. The k required when using the PZFC also increased as the phase-locking cutoff frequency decreased. However, the relative increase in k associated with the PZFC (~138 % between 3.0 kHz and 1.2 kHz) is smaller than the relative increase in k associated with instantaneous-compression types (~200 % on average between 3.0 kHz and 1.2 kHz). At the more realistic phase-locking cutoff frequency of 1.2 kHz, the value of k derived from the instantaneous compression is almost twice that derived from the PZFC.

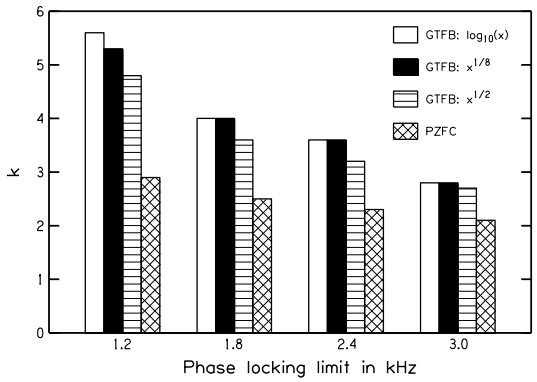

FIG. 9. Bar plot of the expansiveness (in terms of k) of the nonlinearity required to map the perceptual internal h1 to the stimulus h1, displayed as a function of phase-locking filter cutoff frequency for each simulated compression scheme.

V. DISCUSSION

Chapters 1 and 2 showed that pitch-integration time constants depend on the pitch value of the stimuli. Furthermore, comparison of results from Chapters 1 and 2 showed that integration-window time constants depend on the frequency region in which the stimuli are presented. Results from the current study showed that the pitch-integration time constants do not depend on the subjective pitch strength of the stimuli. Taken together, results from the first three chapters suggest that these initial time constants are not

dynamic according to the $H1_S$ of the stimulus but are "hard wired" and scale with the autocorrelation lag, as determined by the scalar, η, where η is dependent on the frequency band in which the stimuli are presented.

In the current study, the raw data suggests that listeners are more sensitive to modulations in temporal regularity when the IRN stimuli have greater n. However, the sensitivity effects could be accounted for once the thresholds were converted into $E(h1_S)$ units using a value of k = 1.2. This value of k was very similar to the value found to relate the pitch strength of RN to RPN stimuli (Wiegrebe et al., 1998).

Chapters 1 and 2 demonstrated that TMTFs were a more informative method of quantifying the temporal resolution of pitch perception because time constants could be estimated directly from the threshold patterns. However, the band-pass characteristic observed in the TMTFs at the slowest modulation rates meant that factors other than sensitivity were responsible for limiting listener performance. This issue became particularly apparent in the current study when using IRNs with low n. In contrast, the gap-detection data did not reach asymptote, even at the longest gaps measured, thus making it impossible to estimate the integration time constants from the data. However, the sensitivity differences between gap-detection thresholds associated with IRNs with different n remained reasonably constant at all gap durations, thus providing a reliable sensitivity measure. Therefore, the gap-detection paradigm is more appropriate than the TMTF paradigm when quantifying differences in sensitivity to modulations in $h1_S$ when pitch strength is low.

The pitch strength that listeners associate with IRNs is mediated by $E(H1_S)$(Yost, 1996). Similarly, data from the current study showed that listeners' sensitivity to modulations in $h1_S$ over time is mediated by the instantaneous $E(h1_S)$. However, nonlinear processes in the auditory periphery compress the relationship between $H1_S$ and $H1_{NAP}$.

Therefore, an expansion using a value of k>1.2 must be applied to $H1_{NAP}$ in order to match $E(H1_S)$.

In the simulations presented in the current chapter, the logarithmic and $x^{1/8}$ compression schemes had almost identical compression ratios in response to a sinusoid. The similarity between the I/O functions relating $H1_{NAP}$ to $H1_S$ when using $x^{1/8}$ and logarithmic compression schemes suggests that cochlea models with similar compression ratios have similar effects on the $H1_{NAP}$ of an IRN. The $x^{1/2}$ compression scheme had a lower compression ratio than the $x^{1/8}$ and logarithmic compression schemes, and the I/O function relating $H1_{NAP}$ to $H1_S$ generated when using the $x^{1/2}$ compression scheme was less compressive than the I/O functions generated from the other instantaneous-compression types. The $x^{1/2}$ and PZFC compression schemes had almost identical compression ratios in response to a sinusoid. However, the I/O function relating $H1_{NAP}$ to $H1_S$ generated when using the PZFC was much less compressive than that generated when using the $x^{1/2}$ compression. If the PZFC and $x^{1/2}$ compression schemes had equal compression ratios in response to IRN stimuli, then one would have expected the I/O functions relating $H1_{NAP}$ to $H1_S$ to be similar for both compression schemes. This suggests that the PZFC was less compressive than the $x^{1/2}$ compression scheme in response to an IRN, even though their compression ratios in response to a sinusoid were similar.

One of the main differences between the PZFC and the gammatone filter bank with instantaneous compression is that the gain applied to the signal by the PZFC is temporally dynamic. The effects of this are illustrated in Fig. 10. Here, the half-wave rectified, compressed output of the channel centred closest to 1 kHz is shown in response to a sinusoid of the same frequency for both the PZFC and gammatone filter banks. The initial build-up of energy is visible in both outputs. The energy within the instantaneously compressed gammatone filter-bank channel increases to a maximum and then remains

constant. The output of the PZFC increases to a maximum, after which the effects of the dynamic gain become apparent as the response drops to a relatively constant level after approximately 10 ms. The temporally dynamic gain applied by the PZFC may result in subtle differences between the compression ratios in response to a sinusoid and an IRN, as the gain applied by the PZFC would vary over time in response to the random variations in the energy spectrum of an IRN stimulus. Fig. 10 also shows a slight phase delay of the PZFC output relative to the GTFB output. The GTFB is arranged in parallel, where each channel is independent from the next. Conversely, the PZFC is modelled as a cascade of filters, the output of which can be extracted at the desired channel. The cascading of filters introduces a delay that is more closely related to the underlying travelling-wave hydrodynamics. However, phase differences between channels do not affect autocorrelation-based pitch-strength model predictions, as phase information is discarded by the autocorrelation process within each channel before the results are summed across channels.

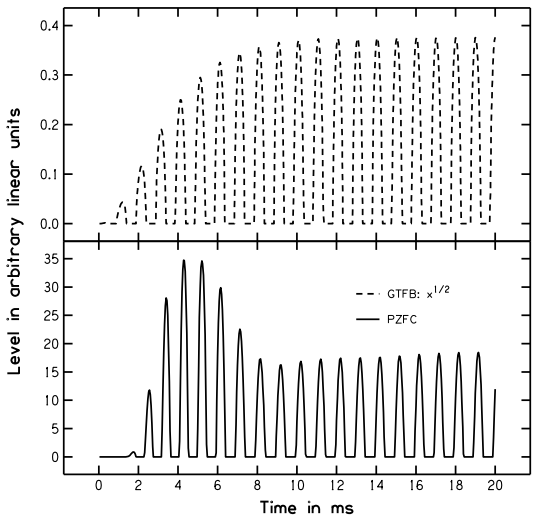

FIG. 10. Output of the channel with a best frequency of 1007 Hz in response to a sinusoidal stimulus of the same frequency, as a function of time. The upper panel shows

the half-wave rectified output of the gammatone filter bank with power-law compression applied. The lower panel shows the output of the PZFC.

The PZFC is also able to account for two-tone suppression data (Sachs and Kiang, 1968). Two-tone suppression describes the phenomenon where an off-frequency stimulus is able to suppress the on-frequency response of a neuron. To account for this, the adaptive gain control (AGC) of the PZFC is not only dependent on the temporal distribution of energy, but also on the spectral distribution of energy across channels. Therefore, unlike the GTFB with instantaneous compression, the compression ratio associated with an individual channel of the PZFC is reduced by energy in off-frequency channels. Therefore, its compression ratio in response to a wide-band stimulus such as an IRN is less than its compression ratio in response to a sinusoid.

Data from the current study suggests that the neural decision mechanism is based on an expanded representation of $H1_S$. However, the relationship between $H1_S$ and $H1_{NAP}$ is compressive, irrespective of the cochlear compression scheme used. Therefore, if an autocorrelation-based pitch-extraction mechanism is responsible for the data measured in the current study, then an expansive mechanism is also required that is likely to have a neural basis. In terms of Licklider's (1951) neural model of pitch extraction, the expansive function could be implemented as an additional neural layer between the coincidence-detection and the leaky-integration layers. Unlike the neural time constants associated with pitch perception, the proposed expansive process does not appear to have any parametric dependencies on the autocorrelation lag, as data from Chapters 1 and 2 showed that listeners are equally sensitive to modulations in $h1_S$, irrespective of the IRN rate. However, the interaction of the phase-locking filter and compression (Fig. 8) suggests that

the proposed expansive mechanism may be dependent on absolute frequency, in that k is likely to be greater in higher-frequency channels.

Chapter 4

Disparity between integration times inferred from the effects of stimulus duration measured in pitch-strength and pitch-value discrimination experiments.

I. INTRODUCTION

As with other senses, in audition, detection and discrimination performance generally improve with increases in stimulus duration (for review, see Viemeister and Plack, 1993). This improvement is referred to as the duration effect. The experiments presented in the current study investigated the integration of pitch cues. It has been shown that pure-tone frequency-discrimination thresholds improve with stimulus duration (Moore 73, Goldstein and Sruvolics 77). Similarly, pitch-discrimination thresholds based on the residue pitch (Schouten et al., 1962) associated with band-limited harmonic complex tones (HCTs) improve with stimulus duration (Plack and Carlyon, 1995, White and Plack, 1998). While performance generally improves with duration, these studies have shown that there is a limit to the duration effect on pitch-value resolution. The underlying assumption has been that pitch-discrimination performance improves with increasing stimulus duration until the pitch-processing mechanism has reached its integration capacity. When the stimulus duration is equal to the integration capacity of the auditory system, then the system cannot accept any further information to improve performance; therefore, performance reaches an asymptote.

Under this assumption, the stimulus duration at which the thresholds reach asymptote has been used to make inferences about the integration time of the system. Results from earlier studies that investigated the duration effect on rate discriminability in HCTs have suggested that the system uses longer integration times for unresolved tonal stimuli compared to resolved tonal stimuli (Plack and Carlyon, 1995, White and Plack, 1998). Later studies however, showed that pitch-value discrimination thresholds reach asymptote at approximately the same critical point when stimulus durations are defined according to the number of cycles of the stimulus waveform (Krumbholz et al., 2003b, White and Plack, 2003), suggesting that pitch-value discrimination performance may also

be dependent on the number of available waveform cycles of the stimulus, rather than absolute stimulus duration.

The temporal resolution of pitch extraction refers to the minimum time interval within which different acoustic events can be distinguished. This minimum time interval is limited by temporal integration, which functions as a moving average filter, reducing the contrast between events on which an outcome can be determined by a decision mechanism. The longer the integration window, the more it attenuates rapid fluctuations in the pitch information. While there have been relatively few studies that have directly measured the temporal resolution of pitch extraction (Wiegrebe, 2001, Chapters 1 - 3), data from each of these studies suggest that the time constant of the integration window scales according to the repetition rate of the stimulus.

The fact that results from both pitch-resolution and more recent pitch-integration studies both suggest that integration times scale according to the stimulus rate may indicate that they reflect a common integration process. This hypothesis is supported by the fact that the integration times derived from the data of Krumbholz et al. (2003b) and Chapter 1 were very similar. Krumbholz et al. (2003b) showed that for stimuli band-limited between 0.8 and 3.2 kHz, pitch value-discrimination thresholds reached asymptote at stimulus durations between ~4 and 8 stimulus cycles. In Chapter 1, when stimuli were presented in a similar spectral region (0.78 - 2.98 kHz), the time constants derived from pitch-strength TMTF measurements were 5.44 stimulus cycles, thus falling directly into the range suggested by the data of Krumbholz et al. (2003b). However, this similarity breaks down in higher frequency regions, as the data presented in Chapter 2 suggested that resolution time constants increase sharply with increasing listening region. In contrast, White and Plack's (2003) data suggests that integration times do not change much with frequency region. They showed that pitch value-discrimination thresholds reached

asymptote at stimulus durations of ~10 stimulus cycles when stimuli were presented in a similar band (2.75 and 3.75 kHz) to that used in Chapter 2. Furthermore, when stimuli were presented between 5.5 and 7.5 kHz, they showed no influence of listening region on the duration at which thresholds reached asymptote, whereas data from the 2nd part of Chapter 2 suggested that the resolution time constants continued to increase with increasing listening region up to at least 4.5 kHz. These results seem to suggest that the pitch-related time constants responsible for limiting temporal resolution are longer than those used by the system when integrating information in order to improve discrimination performance. The current study was aimed at investigating this seeming paradox.

The integration data of Krumbholz et al. (2003b) and the resolution data measured in Chapter 1 both predicted similar integration times. However, both of these studies used iterated rippled noise (IRN) stimuli. In higher frequency regions, pitch-integration studies have generally used HCTs, whereas pitch-resolution studies have used tonal stimuli derived from noise. The differences observed between integration windows measured in integration and resolution studies in higher frequency regions may have been due to differences in the stimuli used. In the first part of the current study, pitch-value discrimination thresholds were measured as a function of the stimulus duration using IRN stimuli. The stimuli and experimental parameters were matched as closely as possible to those used in the resolution experiments presented in Chapters 1, 2, and 3, allowing for direct comparison of results from both integration and resolution paradigms using IRN stimuli. A second experiment was conducted using a similar procedure as the first, but where pitch-strength discrimination thresholds were measured as a function of the stimulus duration. If integration time is reflected by the stimulus duration at which thresholds reach asymptote, then one would expect to see the point of asymptote occur

after the same number of stimulus cycles, irrespective of whether pitch-value or pitch-strength is being discriminated.

II. EXPERIMENT 1: THE DURATION EFFECT FOR PITCH-VALUE DISCRIMINATION

A. Experiment 1a: Parametric effects of repetition rate and listening region

1. Stimuli

IRNs were generated with 16 iterations of the add-original, delay-and-add algorithm (Yost, 1996). The IRNs were generated in the spectral domain to avoid being limited to only using delays at integer multiples of the digital sampling period (Krumbholz et al., 2003a). This was achieved by multiplying the Fourier spectrum of a Gaussian noise with the comb-filter transfer function, H(ω) (as defined by Eqn. 1), of an add-original IRN with delay d and n iterations, where j is the imaginary unit, and ω is angular frequency.

$$H(\omega) = \sum_{k=0}^{n} g^{k} e^{-jk\omega d} \qquad \text{(EQN. 1.)}$$

The gain, g, was always 1 in the current experiments. Stimuli were generated using Fast Fourier Transforms (FFTs) with a minimum 2^{15} points to obtain the desired frequency resolution. Stimuli were subsequently truncated to the desired duration.

To assess the effect of listening region, IRNs were bandpass filtered into either a low-frequency region as described in Chapter 1, or a high-frequency region as described in Chapter 2. The low-frequency cutoff of the low-frequency region was 0.78 kHz, which is within the putative phase-locking range of human inner hair cells. The low-frequency cutoff of the high-frequency region was 2.64 kHz. While the phase-locking limit is not known in humans, the fidelity of the TFS information available in the high-frequency region would be expected to be severely degraded relative to that in the low-frequency

region. To investigate the potential interaction between stimulus repetition rate and the duration effect, four different stimulus repetition rates were used in each band. The repetition rates used here were the same as those used in Chapters 1 and 2 so that direct comparisons could be made between results from the different paradigms. According to Chapter 1 and 2, the higher-rate IRNs in each band contained some resolved harmonics, while the lower-rate IRNs were completely unresolved.

The loudness of a sound with constant intensity is known to increase with increasing duration (Florentine et al., 1993). To compensate for this, shorter stimuli must be presented at a higher level than longer stimuli to achieve an equal loudness percept. Stimuli with durations greater than 100 ms were presented at a level of 60-dB SPL. Stimuli shorter than 100 ms were presented at an increase level relative to their duration to maintain constant energy.

Stimuli were gated on and off with 2.5-ms cosine-squared ramps and were presented in a continuous noise to mask audible distortion products below the stimulus passband. This noise was lowpass filtered at 0.5 octaves below the lower cutoff frequency of the stimulus passband using an 8^{th}-order Butterworth filter. Prior to lowpass filtering, the noise was filtered in the spectral domain so as to produce a roughly constant excitation level of 55 dB SPL per ERB. Stimuli were presented to the listeners using the same equipment described in Chapter 1.

2. Procedure

Each trial consisted of three observation intervals, which were separated by 500-ms gaps. Two intervals contained lower-pitched stimuli while the remaining interval contained the higher-pitched target stimulus. Intervals were presented in a random order within each trial, and the listeners' task was to identify the interval containing the stimulus

with the highest pitch. An adaptive staircase technique was used to measure thresholds, where the adaptive parameter was the difference in the repetition rate between intervals spanning the nominal rate, f_R. The repetition rate differences were quantified using a logarithmic unit of measure, cents, where an equally tempered semitone is equal to 100 cents. Therefore, one cent is equal to the ratio $2^{1/1200}$. Between each trial, f_R was randomly roved (flat distribution) by +/- 50 cents. This was to prevent listeners from basing their decisions on the pitch of stimuli across trials. At the beginning of each threshold run, the adaptive parameter was 600 cents (half an octave), which was well above the anticipated threshold. The adaptive parameter was decreased after two consecutive correct responses and increased after each incorrect response to track the rate difference that yielded 70.7% correct responses (Levitt, 1971). The step size for the increments and decrements in the IRN rate difference was by multiplication and division with a factor of 2 for the first reversal in level, 1.5 for the second reversal, and 1.25 for the rest of the eight reversals that made up each threshold run. The geometric mean of the last six reversals was taken as the threshold estimate for each run.

Thresholds were measured for stimulus durations equal to multiples of the central IRN delay, d. Thresholds were measured for stimulus durations of 4d, 6d, 8d, 16d, and 32d. For the highest IRN rates (300.00 and 424.26 Hz), the 4d condition was too short in terms of absolute stimulus duration for listeners to perform the task. Therefore, the range of durations at which thresholds were measured was 6d, 8d, 12d, 16d, and 32d.

Three threshold runs were conducted for each participant per stimulus condition. Threshold runs were conducted in a random order for each participant until one run of each condition was completed. This process was repeated for the 2nd and 3rd runs of each condition to minimize training effects.

3. Listeners

A total of 4 listeners (3 male and 1 female, aged between 23 and 28 years) participated in the experiment, one of whom was the author. Participants were paid for their services at an hourly rate and met the same criteria outlined in Chapter 1.

4. Results and interim discussion

Data from the current experiment are presented in Fig. 1. The statistical significance of the observations was tested by performing a linear mixed-models analysis on the data. The analysis was performed on factors normalized stimulus duration (duration/d), frequency region, IRN rate, and resolvability. The dependent variable was mean threshold averaged across the three runs for each participant and condition.

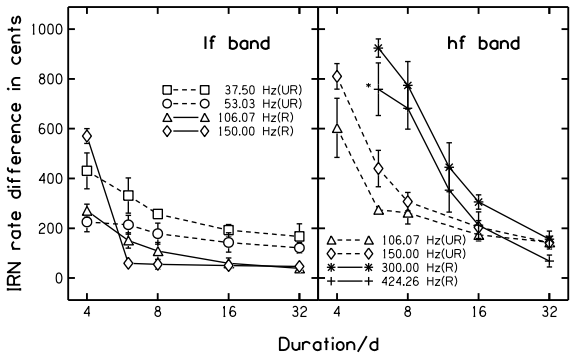

FIG. 1. Pitch-discrimination thresholds plotted as a function of normalized stimulus duration. Mean thresholds are plotted for each condition, averaged across the 4 listeners. Error bars represent the inter-listener standard error. The asterisk identifies a condition where one of the listeners was unable to obtain a threshold. The left-hand panel shows thresholds measured in the low-frequency region, and the right-hand panel shows thresholds measured in the high-frequency region. The parameter is the central reference IRN rate, where dashed lines represent unresolved stimuli and solid lines represent

resolved stimuli. The open triangle and diamond markers are common to both panels and represent the 106.07- and 150.00-Hz conditions respectively.

Generally, listener performance improved with increasing stimulus duration, as shown by the significant main effect of normalized duration in both low- [$F(4,57)=50.775$, $p<0.001$] and high-frequency regions [$F(5,56.026)=42.522$, $p<0.001$]. In the low-frequency region, thresholds for the resolved IRNs were lower overall than those for the unresolved IRNs [$F(1,57)=36.039$, $p<0.001$], and the final asymptotic thresholds for the resolved IRNs were, likewise, lower than for the unresolved IRNs as shown by the pairwise comparison between resolved and unresolved thresholds at the longest duration measured [$F(1,57)=10.233$, $p=0.002$]. Conversely, in the high-frequency region, thresholds for the resolved IRNs were higher overall than those for the unresolved IRNs [$F(1,56.031)=19.581$, $p<0.001$]. This was mainly due to the resolved thresholds being higher than the unresolved thresholds at short normalized duration, rather than due to differences in asymptotic performance at long duration.

In the low-frequency region, thresholds appeared to reach asymptote at approximately the same duration/d, irrespective of the IRN rate. Pairwise comparisons between thresholds at successive stimulus durations were made for each IRN rate. These comparisons generally showed that the duration effect was no longer significant by 6d, suggesting that thresholds reached asymptote somewhere between 6d and 8d.

In the high-frequency region, the unresolved IRNs appeared to reach asymptote at a shorter duration/d than the resolved data. However, this is more likely to reflect limitations associated with the absolute stimulus durations rather than a resolvability dependent difference in the time constants. The value of d is the inverse of the IRN rate; therefore, the value of d associated with the 106.07-Hz IRN is 4 times longer than that

associated with the 424.26-Hz IRN. In the high-frequency region, the absolute stimulus duration of the 424.26-Hz IRN data at duration/d=6 was so short that one listener was unable to perform the task at all. This is denoted by an asterisk in the figure adjacent to the data point in question.

IRNs are made from noise, and the variability in the spectral composition of short noise samples is greater than that in relatively longer noise samples. Fig. 2 shows the spectra of the noises used to make the high-frequency region 106.07- and a 424.26-Hz IRNs with relative stimulus duration of 8d. In absolute terms, the duration of the 106.07-Hz IRN is 75.4 ms, and the duration of the 424.26-Hz IRN is 18.9 ms. The spectra are shown before and after filtering with the IRN transfer function. The short absolute duration associated with the 424.26-Hz IRN gives the noise source a highly variable spectrum compared to that of the 106.07-Hz IRN. When the noises are filtered to make IRN stimuli, the resulting spectrum of the longer stimulus is far more representative of the IRN transfer function than that of the shorter stimulus. Therefore, the difference in duration effects observed between resolved and unresolved thresholds in the high frequency band is more likely to be an artefact of the experimental procedure (relating to variability in the spectral composition of the stimuli) than a resolvability dependent difference in integrations time. Hence, the resolved thresholds in the high frequency band cannot be used to make inferences about integration time and are not considered further. In the low-frequency region, the stimulus-related spectral variability is also likely to explain the relatively high threshold measured for the 150.00-Hz IRN at a duration of 4d (27 ms).

For the lower-rate (unresolved) IRNs in the high-frequency region, pairwise comparisons of successive stimulus durations revealed that the duration effect was no longer significant by 8d. This suggests that the asymptote occurred somewhere between 8d

and 16d in the high frequency band, whereas the asymptote occurred between 6d and 8d in the low-frequency band.

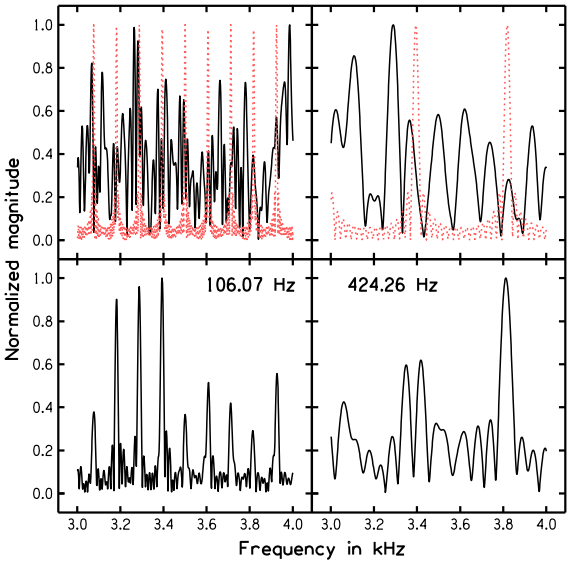

FIG. 2. The upper panels show the spectra of a single noise source that is truncated to a duration of 8d, where d = 1/106.07 on the left-hand side and 1/424.26 on the right-hand side. Superimposed upon the noise spectra are the IRN transfer functions of the 106.07- and 424.26-Hz IRNs used in the current experiment, where n = 16. The resulting IRN spectra are shown in the lower panels.

The finding that thresholds in the low-frequency region had reached asymptote by approximately 6d to 8d was in good agreement with the findings of Krumbholz et al. (2003b), where thresholds reached asymptote by approximately 4d to 8d for similarly filtered IRN stimuli. Furthermore, the asymptote in the duration effect measured in the high-frequency region of the current study was in fairly close agreement with the findings of White and Plack (2003), who measured the duration effect on pitch discrimination thresholds using HCTs similar frequency region. However, if one were to infer integration

times from the duration at which pitch discrimination thresholds reached asymptote, then data from the current study suggests that integration times are shorter than those derived to account for the resolution data presented in Chapters 1 and 2. This paradoxical result is addressed in section IV.

B. Experiment 1b: Parametric effect of n

1. Methods

The goal of this experiment was to measure the duration effect in the same parameter space as the resolution experiment presented in Chapter 3. In this experiment, pitch-discrimination thresholds were again measured as a function of stimulus duration using the same experimental procedure outlined in experiment 1a. Here, the main experimental parameter was the number of iterations, n, used in the IRN circuit. As in Chapter 3, thresholds were measured for n = 1, 2, 4, and 8 at each stimulus duration. IRNs were also presented at the same rates and in the same spectral band used in Chapter 3: IRNs were filtered into a band between 0.78 and 2.98 kHz and thresholds were measured around two nominal IRN rates, including 53.03 Hz (unresolved) and 106.07 Hz (resolved). Four listeners took part (2 male, 2 female, aged between 21 and 26), one of whom was the author. Listeners met the same criteria outlined in experiment 1a.

2. Results and interim discussion

Data from the current experiment are presented in Fig. 3. The statistical significance of the observations was tested by using a similar analysis to that used in experiment 1a, but with n as an additional factor. There was a clear overall duration effect, as shown by the significant main effect of duration/d [$F(4,117) = 138.739$, $p < 0.001$]. There was also significant overall main effect of n [$F(3,117) = 16.757$, $p < 0.001$], due to

the fact that thresholds were slightly higher on average for stimuli with lower n at durations shorter than the point of asymptote. However, the final asymptotic thresholds were all very similar as shown by the insignificant pairwise comparisons of thresholds for both resolved [$F(3,117)=0.002$, $p>0.999$] and unresolved [$F(3,117)=0.604$, $p=0.614$] stimuli at the longest durations measured.

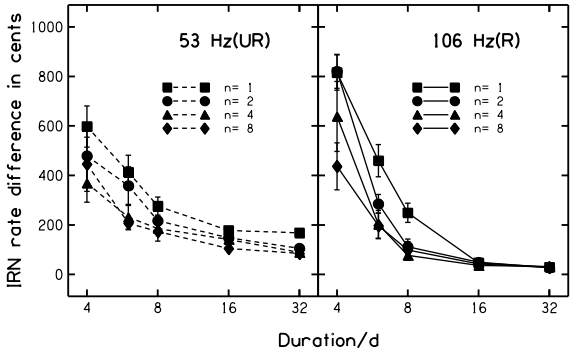

FIG. 3. Pitch-discrimination thresholds plotted as a function of stimulus duration, normalized by the IRN rate. Mean thresholds are plotted for each condition, averaged across the 4 listeners. Error bars represent the inter-listener standard error. The left-hand panel shows thresholds that were measured, centred around an unresolved rate of 53.03 Hz, while the right-hand panel shows thresholds that were measured, centred around a resolved rate of 106.07 Hz. The parameter is n. As in Fig. 1, dashed lines represent unresolved stimuli and solid lines represent resolved stimuli.

No significant effect of IRN rate was observed [$F(1,117)=0.849$, $p=0.359$]. However, there was a significant interaction of IRN rate and duration/d [$F(4,117)=138.739$, $p<0.001$]. This suggested that thresholds for the different IRN rates reached asymptote at different values of duration/d. However, pairwise comparisons between thresholds at successive durations for each value of n revealed that, in general,

the duration effect was no longer significant after 8d. This was with the exception of the n = 1 condition in the 106.07-Hz IRN, where the duration effect was significant up to 16d. However, this does not necessarily suggest that the system is integrating over a longer window for this one set of stimulus parameters. The absolute stimulus durations associated with the stimuli centred around 106.07 Hz were half the length of those associated with the stimuli centred around 53.03 Hz. Therefore, one would expect twice the spectral variability in the stimuli centred around 106.07 Hz. Furthermore, the differences between the spectral variability associated with the 53.03- and 106.07-Hz stimuli would be exacerbated in the n=1 conditions due to the relatively broad peaks of the n=1 IRN transfer functions.

III. EXPERIMENT 2: THE DURATION EFFECT IN PITCH-STRENGTH DISCRIMINATION

A. Methods

1. Stimuli

In the current experiment, the duration effect was measured in a pitch-strength discrimination task. The pitch strength associated with a stimulus is proportional to the amount of temporal regularity within the stimulus. At one extreme is Gaussian noise, which has no temporal regularity and thus has no associated pitch. At the other extreme is a periodic stimulus, which is deterministic and thus gives rise to a clearly tonal percept. In previous chapters, the temporal regularity within IRNs was changed over time by adjusting the gain parameter, g, in the dynamic IRN circuit introduced in Chapter 1. In each iteration of the circuit, g controls the mix ratio of the delayed IRN signal with an uncorrelated noise. The pitch strength of an IRN stimulus is monotonically related to the height of the peak ($H1_S$) occurring at a lag equal to d in the autocorrelogram of the

stimulus. In the dynamic IRN circuit, the relationship between g and $H1_S$ is defined by Eqn. 2.

$$H1_S = \frac{n}{n+1} \cdot g \qquad \text{(EQN. 2.)}$$

As in Experiment 1b, IRNs were generated in the frequency domain at rates of 53.03 Hz (unresolved) and 106.07 Hz (resolved), then filtered into the same listening region (0.78 – 2.98 kHz). IRNs were generated using n=16, and stimuli were presented at the same levels described in Experiment 1a.

2. Procedure

Pitch-strength discrimination thresholds were measured at normalized stimulus durations of 4d, 8d, 16d, 32d, 64d, and 128d for the 53.03-Hz conditions and 4d, 8d, 16d, 32d, 64d, 128d, and 256d for the 106.07-Hz conditions. The same adaptive staircase technique used in Experiments 1a and 1b was used again here. Each trial consisted of three observation intervals that were separated by 500-ms gaps. The task was to detect the interval containing the stimulus with different pitch strength to the other two intervals.

In order to make results comparable to interaural correlation discrimination tasks measured in the binaural domain (Pollack and Trittipoe, 1959), thresholds were measured for the smallest detectable increase in g from a reference g=0, and also for the smallest detectable increase from a reference g=1. To simplify the experimental procedure, g was adjusted by controlling the mix ratio (MR) of the frequency-domain-generated IRN with an uncorrelated noise source, where MR was the adaptive parameter in the tracking process. The relationship between MR and g is defined by Eqn. 3. Therefore, $H1_S$ can be calculated from MR by substituting Eqn. 3. into Eqn. 2.

$$g = \frac{MR}{MR+1} \qquad \text{(EQN. 3.)}$$

For the task where listeners had to detect a reduction in g from a reference g=1, two observation intervals contained IRNs generated from independent noise sources, and the remaining interval contained an IRN mixed with noise. The adaptive parameter was MR, which was increased after two consecutive correct responses and decreased after each incorrect response. The step size for the increments and decrements in the adaptive parameter was 5 dB for the first reversal, 2 dB for the second reversal, and 1 dB for the rest of the eight reversals that made up each threshold run. For the task where listeners had to detect an increase in g from a reference g=0, two observation intervals contained Gaussian noises and the remaining interval contained an IRN mixed with noise. The adaptive procedure was simply reversed, so MR was decreased after two consecutive correct responses and increased after each incorrect response.

3. Listeners

Five listeners took part (2 male, 3 female, aged between 21 and 37), one of whom was the author. Listeners met the same criteria outlined in Experiment 1a.

B. Results and interim discussion

Data from the current experiment are presented in Fig. 4. The statistical significance of the observations was tested by performing a linear mixed-models analysis on the data. The analysis was performed on factors resolvability, normalized stimulus duration (duration/d), and task. The dependent variable was mean threshold averaged across the three runs for each of the 5 participants per condition.

Listeners' performance in the current pitch-strength discrimination experiment improved with increasing stimulus duration. This was shown by the significant overall main effect of duration/d [$F(6,100)=79.076$, $p<0.001$]. At all stimulus durations, listeners were much more sensitive to reductions in pitch strength (from g=1) than to increases in pitch strength (from g=0). This was shown by the significant effect of task [$F(1,100)=754.970$, $p<0.001$]. In the binaural domain, Pollack and Trittipoe (1959) were the first to show that the change in interaural correlation (squared) required for 75% correct identification varied from 0.44 for a reference correlation of 0 to approximately 0.04 for a reference correlation of 1. The similarity in the asymmetries in thresholds observed in the pitch-domain and the binaural-domain data may suggest that very similar mechanisms may be responsible for extracting interaural cross-correlation and monaural serial correlation.

In the data presented in the current study, there was no main effect of resolvability or interaction between task and resolvability; however, there was a significant interaction of resolvability and normalized stimulus duration [$F(5,100) = 16.027$, $p < 0.001$]. This was likely brought into significance by the relatively high thresholds of the 106.07-Hz IRNs in both tasks when duration=4d. As observed in the pitch-value discrimination data, these outlier thresholds were probably the result of procedural limitations associated with high stimulus-related variability at the very short absolute stimulus durations (37.7 ms) of these conditions.

Stimulus durations of up to 4 times longer than those presented in the pitch-value discrimination experiments were used; nevertheless, the pitch-strength discrimination thresholds presented here did not appear to have reached a clear asymptote, even at the longest stimulus durations measured (2414 ms). Therefore, the integration times reflected by the thresholds measured here only appear to be limited by the stimulus duration and are

far longer than those reflected by the pitch-value discrimination experiments. A linear regression of the mean thresholds for detecting an increase in g – excluding the 106.07-Hz outlier threshold at duration of 4d – gave a slope of -1.49 dB per doubling of stimulus duration with a log-linear intercept of 5.26 dB and a correlation coefficient (r^2) of 0.93. A similar regression performed on the thresholds for detecting a decrease in g, omitting the outlier at duration/d = 4, gave a slope of -1.06 dB per doubling of stimulus duration with an intercept of -5.75 and r^2 of 0.94. There was no significant interaction between normalized duration and task; thus thresholds can be said to decrease at an overall rate of about 1.28 dB per doubling of stimulus duration.

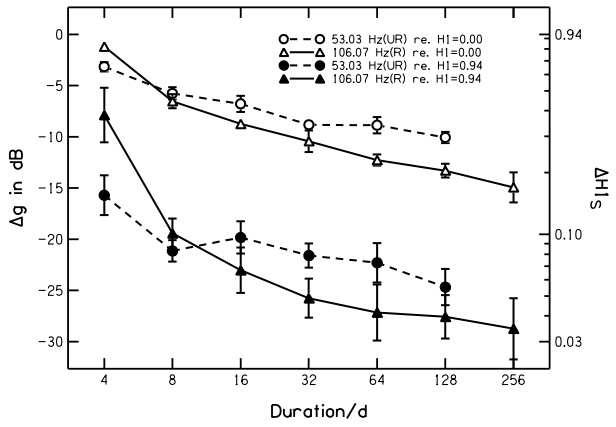

FIG. 4. Pitch-strength discrimination thresholds averaged across listeners. Thresholds are shown in units of the smallest detectable change in g on the left axis and in units of the smallest detectable change in $H1_S$ on the right axis. Filled symbols represent thresholds for detecting a reduction in g, and open symbols represent thresholds for detecting an increase in g. The unresolved 53.03-Hz IRN conditions are denoted by circles connected with dashed lines, and the resolved 106.07-Hz IRN conditions are denoted by triangles connected with solid lines. Error bars represent inter-listener standard error.

IV. MODELLING

A. Comparison of variability in L1 and H1 measurements

Thresholds measured in the pitch-strength discrimination task suggested that integration was only limited by the stimulus duration. In contrast, listeners reached asymptotic performance in the pitch-value discrimination tasks at relatively short stimulus durations. If the duration at which thresholds reached asymptote reflects the integration capacity of the system, then one would have expected to see similar duration effects for both experimental paradigms. However, this was not the case.

The stochastic nature of IRN stimuli means that the measured values of the lag at which the first peak occurs in the autocorrelogram ($L1_S$) and the height of that peak ($H1_S$) are likely to vary between stimuli, and the variation is likely to be bigger for shorter stimuli. If the pitch processor is able to integrate across the entire stimulus, then the amount of variance between measurements would decrease with increasing duration as the noise component of the stimulus is averaged out. Therefore, an alternative hypothesis is that the thresholds reach asymptote at durations by which the variance becomes negligible relative to the resolution with which pitch strength or pitch value can be represented internally. Under this hypothesis the asymptote does not necessarily reflect the integration capacity of the system.

For a given stimulus duration, the relative variability in $H1_S$ and $L1_S$ are likely to be quite different. Furthermore, the rates of decrease in variability with increases in stimulus duration are also likely to be quite different for $L1_S$ compared to $H1_S$. This may be able to explain why listener thresholds reached asymptote by relatively short durations in the pitch-value discrimination tasks, while they did not reach asymptote in the pitch-strength discrimination task at all. This hypothesis was tested by measuring and comparing

the variance in $L1_S$ and $H1_S$ between stimuli over the range of stimulus durations tested experimentally.

For $L1_S$, histograms were generated by measuring the lag in the region of 1d at which the autocorrelation was maximum in response to 1000 IRNs that were 100 cents higher and to another 1000 IRNs that were 100 cents lower than a central value of 106.07 Hz. Separate histograms were generated in response to stimuli with duration/d = 4, 8, 16, and 32. Stimuli were produced as described in Experiment 1b. The interval of +/- 100 cents was chosen, as this was just above the asymptotic thresholds measured in Experiment 1b. At each duration, histograms were generated in response to IRNs with n = 1, 2, 4, 8, and 16. The resolution of the histograms was limited by the sampling period (1/25 kHz).

For $H1_S$, histograms were generated by recording the height of the peak occurring at $L1_S$ in the normalized autocorrelation function in response to 1000 stimuli with n = 16. A single IRN rate of 106.07 Hz was used, as no effects of harmonic resolvability were shown experimentally. Histograms were generated for the stimulus durations tested in Experiment 2, at mix ratios of -∞, -5, 0, 5, and ∞ dB between IRN and Gaussian noise. These mix ratios correspond to fairly evenly distributed g values of 0, 0.24, 0.50, 0.76, and 1.

In order to quantify the effects of the auditory periphery on the variance in the pitch estimates, histograms were also generated in response to the simulated neural activity pattern (NAP). The NAPs were generated using the peripheral model described in Chapter 1, and histograms of $L1_{NAP}$ and $H1_{NAP}$ were generated using the same methods used to generate histograms of $L1_S$ and $H1_S$. The results of the analyses of the signal and NAP are shown adjacent to one another for $L1_S$ and $L1_{NAP}$ in Fig. 5, and for $H1_S$ and $H1_{NAP}$ in Fig. 6.

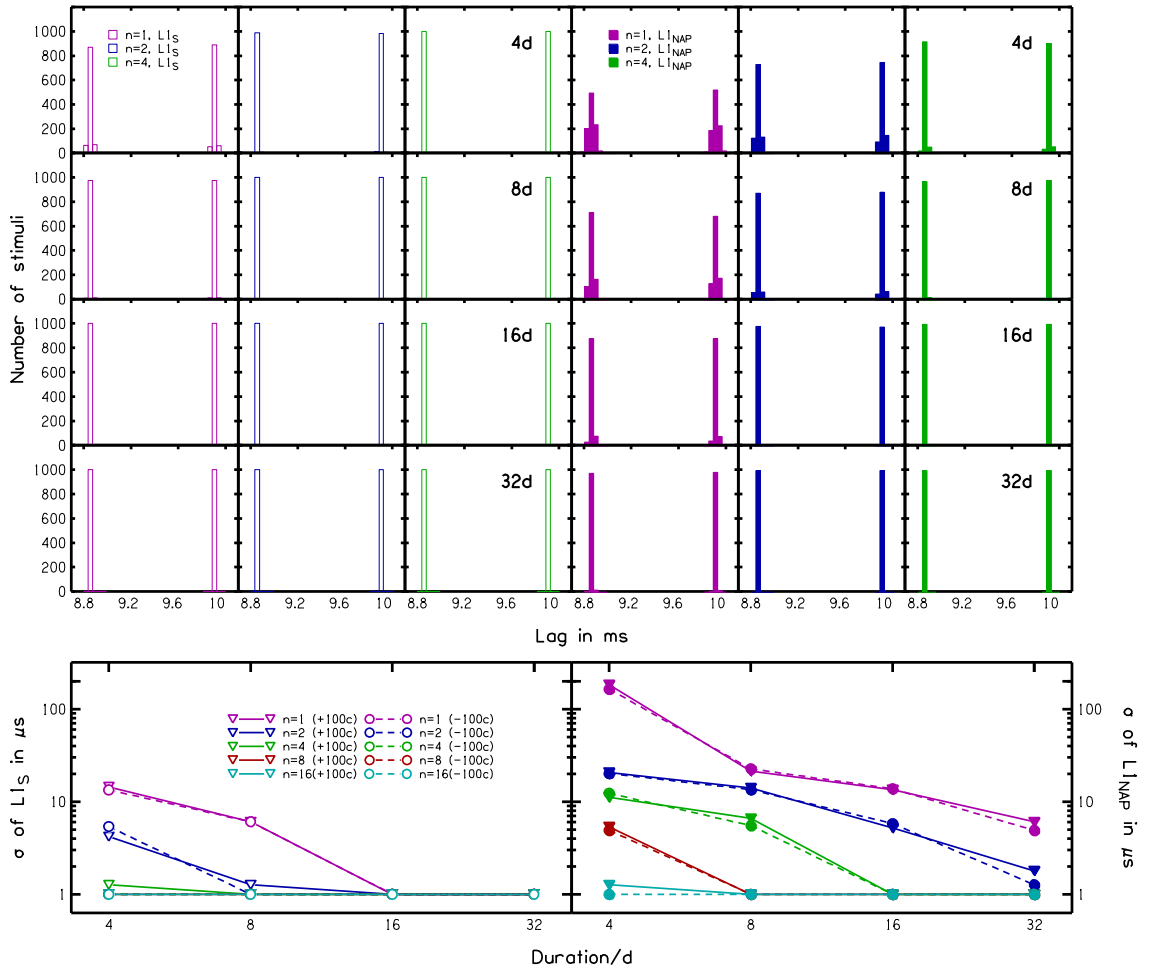

FIG. 5. The left half of the figure shows the results of the analysis of $L1_S$, and the right half of the figure shows the results of the analysis of $L1_{NAP}$. From here on, open symbols and histogram bars represent results of analyses performed directly on the stimuli, whereas closed symbols and histogram bars represent results of analyses performed on NAPs in response to the stimuli. Each column of the upper group of panels shows the distributions of $L1_S$ for IRNs with n=1 (left panel), n=2 (central panel), and n=4 (right panel). Each row of the smaller panels from top to bottom shows histograms generated in response to stimuli that had durations of 4, 8, 16, and 32d. The large panels at the bottom of the figure show the standard deviations of all of the histograms generated (including the n=8 and n=16 histograms not shown in the smaller panels) as a function of stimulus duration.

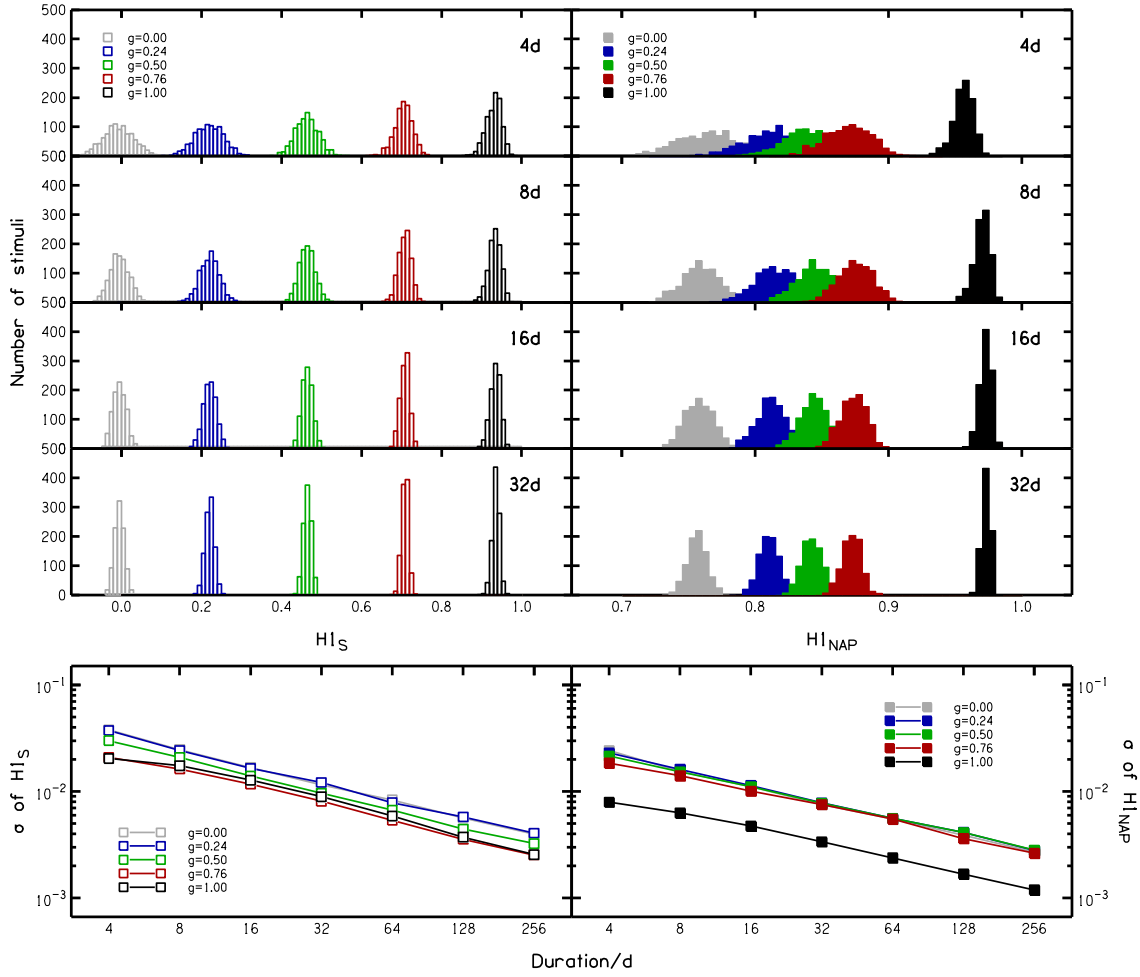

FIG. 6. The left half of the figure shows the results of the analysis of $H1_S$, and the right half of the figure shows the results of the analysis of $H1_{NAP}$. Each column of the smaller panels in the top half of the figure shows histograms in response to stimuli at the various values of g used. The extremes of this range (g=0 and g=1) are shown in monochrome shades. Each row of the smaller panels from top to bottom shows histograms generated in response to stimuli that had durations of 4, 8, 16, and 32d. The lower panels show the standard deviations of all of the histograms generated as a function of stimulus duration. This includes durations up to 256d, for which histograms are not shown in the upper panels.

There was relatively little variance in the $L1_S$ histograms, even at the shortest stimulus durations used. This is given by the fact that the distances between the histograms shown in Fig. 5 are very large in comparison to their widths. The smallest measurable standard deviation, σ_{min}, occurs when just a single stimulus out of N total stimuli has a value different to the mean by an amount equal to a single sampling period, T_S. The standard deviation in this condition is given by Eqn. 3.

$$\sigma_{min} = \frac{T_S}{\sqrt{N}} \qquad \text{(EQN. 3.)}$$

In the current analysis, $T_S = 1/25x10^3$ and N = 1000; therefore, the smallest measurable standard deviation was 1.26 μs. Histograms are only shown for n = 1, 2, and 4, because variance had dropped below σ_{min} by n = 4. The histograms of $L1_S$ also showed that even for the shortest durations and the lowest number of iterations, n, the measured values of $L1_S$ never deviated from the mean by more than a single sampling period. As the stimulus duration was increased, the variability in the distributions decreased. The variance was below σ_{min} at stimulus durations greater than 8d, even for the n = 1 condition. The histograms of $L1_{NAP}$ had considerably more variability than the histograms of $L1_S$ at equal n and duration. However, the distance between the histograms was still large in comparison to their width.

The bottom panels of Fig. 5 show the standard deviation of the histograms of $L1_S$ and $L1_{NAP}$ as a function of the stimulus duration. Standard deviations lower than σ_{min} (1.26 μs) were represented by a value of 1.00 μs in the figure. In general, the standard deviation of both $L1_S$ and $L1_{NAP}$ was very small in comparison to the distance between the distributions, even at the shortest stimulus durations used. Furthermore, the variance dropped to negligible levels after relatively short stimulus durations. The stimuli with lower n generally had a higher standard deviation on average; however, as measured in

Experiment 1b, there was very little effect of n on the durations required for listeners to reach asymptotic thresholds. The fact that variance in $L1_{NAP}$ depended so strongly on n suggests that the duration effect measured experimentally was determined primarily by an additional source of internal variance related to the variability in neural spiking, rather than stimulus-related noise.

The widths of the histograms of $H1_S$ shown in Fig. 6 were slightly larger at lower g. This meant that the addition of Gaussian noise to an IRN not only lowered the mean value of $H1_S$ but also increased the variance of $H1_S$ between stimuli of equal duration. Like $L1_S$, the variance in the $H1_S$ measures decreased with increasing stimulus duration. However, there was still substantial variability in the histograms at duration=32d. Unlike $L1_S$, there was still substantial variance in the $H1_S$ histograms at the longest stimulus durations tested.

The background level of the autocorrelogram in response to the NAP was higher than the background level of the autocorrelogram in response to the stimulus. For a visual comparison, refer to Fig. 7. The reasons for this increase were discussed in detail in Chapters 2 and 3. The increase in background level compressed the dynamic range of $H1_{NAP}$ relative to $H1_S$ by increasing lower values and leaving higher values relatively unaffected. Therefore, the variance in $L1_{NAP}$ was greater than that in $L1_S$ because the noise-induced peaks away from the lag that was equal to d were greatly increased. The effects of the compressed peak-to-background ratio had implications for both the mean and the variance in the $H1_{NAP}$ distributions relative to the $H1_S$ distributions. The most obvious change was that the mean values of the distributions for stimuli with different g were compressed together. The effect was so pronounced that the $H1_{NAP}$ distributions of the stimuli generated with different values of g generally overlapped one another. Furthermore, for stimuli with lower g and hence lower mean $H1_{NAP}$ values, the variance

was reduced relative to $H1_S$ more than for stimuli with higher mix ratios. Looking at the plot of how the standard deviation of the distributions changes as a function of stimulus duration shows that the standard deviations of the distributions are equalized by the compressive effect of the periphery, with the exception of the pure IRN stimulus.

Data from Experiment 2 showed that listeners were more sensitive to reductions in correlation from a highly correlated IRN compared to increases in correlation relative to a noise. The asymmetry in the means of the $H1_{NAP}$ distributions compared to the $H1_S$ distributions may be able to account for the task-dependent differences in sensitivity observed in the data, as the $H1_{NAP}$ distributions at different mix ratios were closer to the noise distribution (g=0) than the IRN distribution (g=1). To test this, an index of detectability was calculated from the distributions.

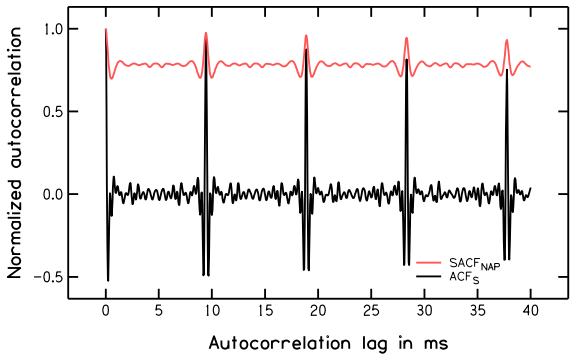

FIG.7. Autocorrelation functions in response to a 106.07-Hz IRN with a duration of 60 seconds. The autocorrelation function in response to the stimulus (ACF_S) is shown by the black line. $H1_S$ and $L1_S$ are derived from this. The summary autocorrelation function in response to the NAP ($SACF_{NAP}$) from the same stimulus is shown by the red line. $H1_{NAP}$ and $L1_{NAP}$ are derived from this.

B. Discriminability based on H1 distributions

Signal detection theory (Green and Swets, 1966) states that the discriminability of two stimuli is inversely proportional to the overlap of the distributions of each stimulus along an internal response axis. If the distributions of each stimulus along the response axis can be estimated, then an index of discriminability can be calculated.

The effective mean, μ_E, is a dimensionless, normalized measure of the mean of a distribution in relation to its variance. It is defined as the ratio of the mean of a distribution to the standard deviation. This provides the experimenter with a single statistic defining a distribution, from which the difference between two distributions can be calculated. The overall standard deviation of two separate Gaussian distributions can be calculated as the orthogonal distance between the two standard deviations. The mean difference between the distributions can be calculated by simply subtracting one from the other. Therefore, the effective mean of the difference distribution of reference and signal distributions, $\Delta\mu_E$, can be calculated from the mean of the reference distribution, μ_R, the mean of the signal distribution, μ_S, the standard deviation of the reference distribution, σ_R, and the standard deviation of the signal distribution, σ_S. This index of discriminability is defined by Eqn. 4.

$$\Delta\mu_E = \frac{\mu_S - \mu_R}{\sqrt{\sigma^2{}_S + \sigma^2{}_R}} \qquad \text{(EQN. 4.)}$$

Initially, $\Delta\mu_E$ was calculated for the measured $H1_S$ distributions. For the task where listeners had to detect an increase in g relative to g=0, the reference distribution was the $H1_S$ distribution corresponding to g=0. Values of $\Delta\mu_E$ were then calculated when the signal distributions were the $H1_S$ distributions corresponding to g= 0.24, 0.50, and 0.76. Therefore, the differences in g relative to the reference (Δg) were also Δg=0.24, 0.50, and 0.76 respectively. For the task where listeners had to detect a decrease in g relative to g=1, the reference distribution was the $H1_S$ distribution corresponding to g=1. Again, values of

$\Delta\mu_E$ were calculated when the signal distributions were the $H1_S$ distributions corresponding to g= 0.24, 0.50, and 0.76. However, because the reference distribution is g=1 in this case, the values of Δg are equal to 1-g.

Values of $\Delta\mu_E$ were calculated as a function of stimulus duration (Fig. 8) for both tasks. Like in the listener data, the calculated values suggest that discriminability in both tasks continually improves as a function of stimulus duration across the entire range of durations tested. The calculated values also suggest that for each of the stimulus durations, reductions in g relative to g=1 were more readily detectable than increases in g relative to g=0. This is due to the larger variance in $H1_S$ distribution associated with g=0 compared to the $H1_S$ distribution associated with g=1. However, the asymmetry is different for each Δg. The asymmetry is sizable for the Δg=0.24 conditions and almost non-existent for the Δg=0.76 conditions. This suggests that if listener thresholds were based on $H1_S$, then thresholds for both tasks would be very similar at short durations and would then diverge at longer durations. However, the thresholds measured in Experiment 2 showed a constant sensitivity difference.

Values of $\Delta\mu_E$ were also calculated from the $H1_{NAP}$ distributions as a function of stimulus duration, as shown by Fig. 9. The compressive effect of the peripheral processing meant that the values of $\Delta\mu_E$ were calculated based on much smaller mean differences between signal and reference distributions. Therefore, the values of $\Delta\mu_E$ calculated from the $H1_{NAP}$ distributions were generally lower than those calculated from the $H1_S$ distributions. Lower values of $H1_{NAP}$ were increased more than higher values; hence there was much more overlap between the signal and the reference distributions in the task for detecting increases in g relative to g=0, when compared to the task for detecting decreases in g relative to g=1. For $H1_S$, the difference between the detection of an increase in

correlation from g=0 and a decrease in correlation from g=1 is dependent on the gain. With $H1_{NAP}$, difference becomes independent of gain, which is more in line with the data. The values of $\Delta\mu_E$ calculated from the $H1_{NAP}$ distributions suggest that a Δg of 0.24 relative to g=1 is almost equally discriminable as a Δg of 0.76 relative to g=0 at all stimulus durations. In terms of g (squared), this model predicts a sensitivity difference of approximately 10 dB between tasks, irrespective of stimulus duration.

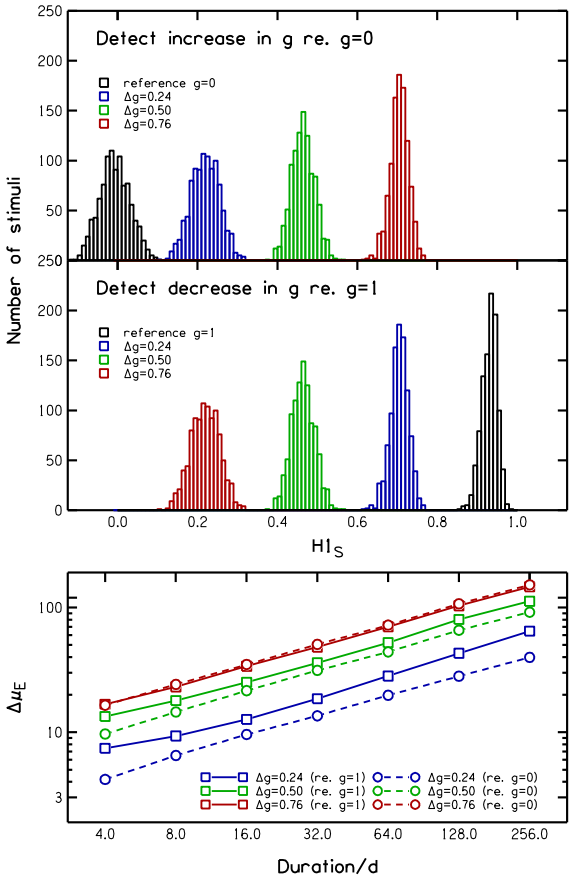

FIG. 8. The top panel shows the distributions used to calculate $\Delta\mu_E$ for the detection of an increase in correlation relative to g=0. The central panel shows the distributions used to calculate $\Delta\mu_E$ for the detection of a reduction in correlation relative to g=1. Distributions were plotted for the shortest stimulus duration (duration=4d) to clearly show the variance differences between the different *Δg histograms*. In each of the upper panels, the black

distribution represents the reference distribution, and the coloured distributions represent the signal distributions. The bottom panel shows $\Delta\mu_E$ for each Δg and each task as a function of stimulus duration.

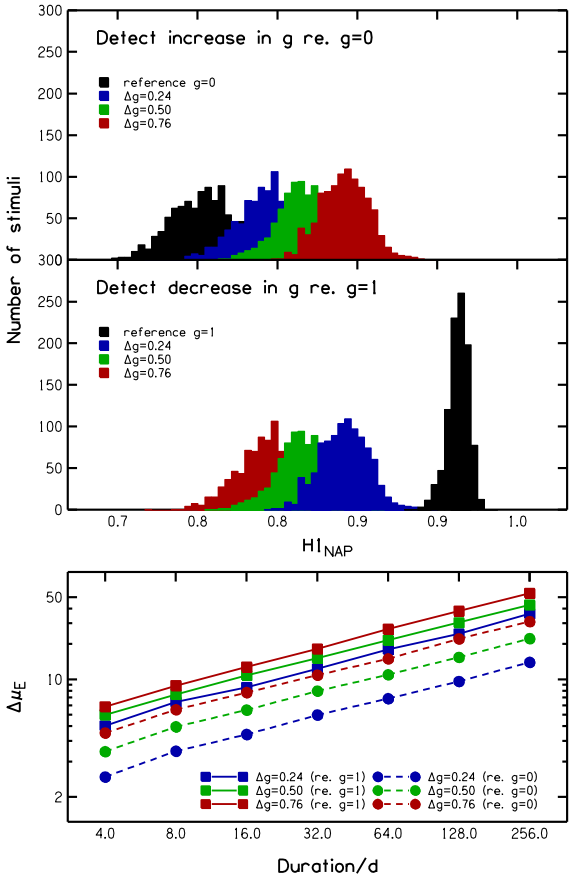

FIG. 9. The data shown in this figure is based on $H1_{NAP}$. The figure is identical in format to Fig. 8.

In Chapter 3, it was shown that that the representations of pitch strength in the central auditory system are likely to be based on an expanded version of $H1_{NAP}$, $E(H1_{NAP})$. In order to see the effects of this expansion on the simulated detectability of changes in pitch strength, values of $\Delta\mu_E$ were calculated from $E(H1_{NAP})$ using the appropriate expansive function (k = 5.6) for the peripheral model used (based on findings from Chapter 3). The results of this analysis are shown in Fig. 10. Looking at the histograms of

$E(H1_{NAP})$, the expansive function restored the overall dynamic range of values that were compressed by the peripheral simulation. However, the expanded distributions are very different to the $H1_S$ distributions shown in Fig. 8. Rather than restore the equal spacing between the different distributions present in $H1_S$, the expansive function exaggerated the asymmetry between the means of the distributions in $H1_{NAP}$. At the same time, the variance of the low g distributions was reduced and the variance of the high g distributions was increased by the expansive process. The values of $\Delta\mu_E$ calculated from the $E(H1_{NAP})$ distributions indicate that the expansive process exaggerates the asymmetry in discriminability between the tasks.

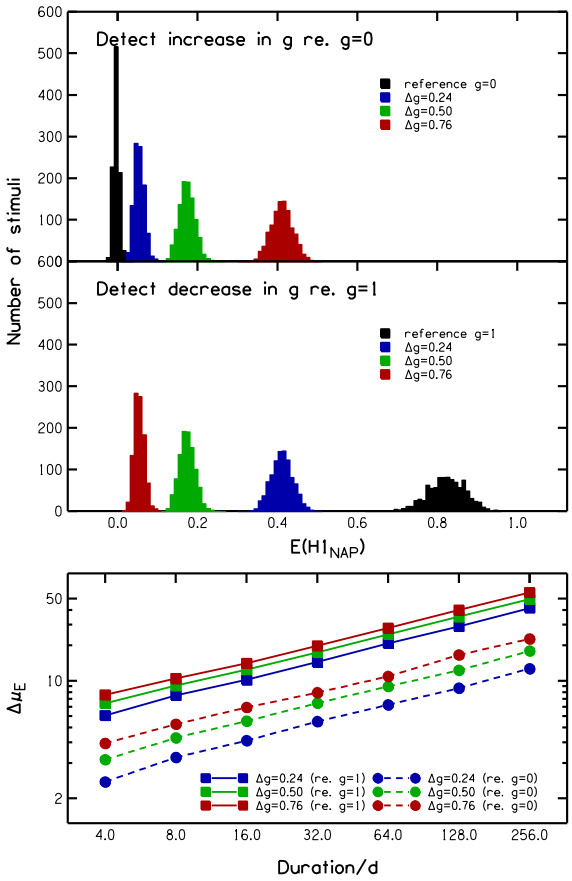

FIG. 10. The data shown in this figure is based on $E(H1_{NAP})$. The figure is identical in format to Figs. 8 and 9.

V. DISCUSSION

Data from Experiment 1 showed that pitch-value discrimination thresholds reached asymptote by relatively short stimulus durations. The duration effect measured in the low-frequency region ceased by approximately 8 stimulus cycles. This was a very similar duration effect to that observed by Krumbholz et al. (2003b). There were some relatively subtle effects of listening region and n, in that thresholds reached asymptote at slightly longer durations when stimuli were presented in a high-frequency region, as also observed by White and Plack (2003), or when stimuli had low n. However, these effects were very small in comparison to the differences observed between the duration effect in pitch-value discrimination and pitch-strength discrimination tasks. Thresholds measured in the pitch-strength discrimination task continued to decrease over the entire range of stimulus durations measured, showing no sign of reaching asymptote. This suggested that pitch integration was being performed over an extremely long duration, greater than 2 seconds. Assuming that pitch value and strength are extracted by a common mechanism, then the data from the pitch-strength discrimination task suggests that something other than the length of the integration window is responsible for limiting performance in the pitch-value discrimination tasks.

In the modelling section, the variance measured in $L1_S$ and $H1_S$ as a function of the stimulus duration qualitatively reflected the listeners' thresholds from the pitch-value and pitch-strength discrimination paradigms respectively. The variance in $L1_S$ was small, even at the shortest durations, and dropped to a negligible level by extremely short stimulus durations. Conversely, the variance in $H1_S$ was large relative to the dynamic range (0 to 1), and the variance continued to decrease at a constant rate over the range of durations measured. This suggested that thresholds reached asymptote when the variance in the pitch measure became negligible, possibly relative to the resolution with which the estimate is

represented internally, and not because the system had reached some sort of integration capacity.

The variance model presented in the current study did not include any stimulus-independent internal noise contributions originating from simulations of stochastic neural processes. If discrimination thresholds were based purely on stimulus-induced noise, then pitch-discrimination thresholds for deterministic stimuli would reach asymptote as soon as the stimulus duration exceeded a single cycle. The variance in $L1_{NAP}$ was greater than that in $L1_S$. Even so, the variance in $L1_{NAP}$ generally decreased to a negligible level by shorter stimulus durations than that required for listeners to reach asymptotic threshold. Furthermore, there was a large effect of n in the simulations and very little effect of n in the data. This suggests that the stimulus-induced variance is negligible in comparison to the stimulus-independent variance in terms of $L1_{NAP}$. The addition of stimulus-independent noise components to the simulations would increase the variance in $L1_{NAP}$. Therefore, greater stimulus durations would be required for the variance in $L1_{NAP}$ to drop to negligible levels. Even so, the stimulus-induced variance measured in the current study was qualitatively able to explain the vast differences between the duration effects observed in the pitch-value compared to the pitch-strength discrimination task.

The means of the $H1_S$ distributions remained constant, irrespective of stimulus duration, whereas the variance decreased by a constant factor per doubling of stimulus duration. Therefore, the calculated discriminability index increased at a rate inversely proportional to the combined variance of the signal and reference $H1_S$ distributions. This constant increase in the detectability index was able to account for the constant reduction in thresholds measured in Experiment 2. However, it was unable to describe the asymmetry in thresholds observed between the tasks of detecting increases and decreases in correlation. The discriminability index calculated from the $H1_{NAP}$ distributions was in

line with the data, suggesting that listeners should be more sensitive to reductions in pitch strength from a highly correlated stimulus compared to increases in pitch strength relative to noise. This asymmetry was primarily due to the shift in the means of the distributions caused by the increased background level of the autocorrelogram of the NAP (Fig.7) relative to the background level of the autocorrelogram of the stimulus. Incidentally, the sensitivity differences in terms of g predicted based on $H1_{NAP}$ were 10 dB between the tasks. This was somewhat less than the difference observed in the data. However, the specific sensitivity difference predicted by the model is dependent on parameters in the peripheral simulation, such as phase locking and compression (for a detailed discussion, refer to Chapter 3). Adjustment of these parameters would greatly change the peak-to-background ratio in the autocorrelogram of the NAP. Therefore, the simulated sensitivity difference between tasks would change accordingly. However, the model does provide a qualitative explanation for the observed sensitivity differences between the tasks in the pitch-strength discrimination experiment.

In the pitch-strength discrimination simulations presented in the current study, the values of $H1_S$ and $H1_{NAP}$ were calculated using an unbiased autocorrelation integrated across the stimulus duration. Therefore, the mean of the distributions remained relatively constant, irrespective of stimulus duration. However, intensity integration studies have shown that the percept of loudness increases with increasing stimulus duration when the RMS level of the stimulus remains constant. Should a similar integration process be responsible for both loudness and pitch integration, then one would expect the percept of pitch strength to also increase with increasing stimulus duration. Furthermore, if pitch-strength integration is based on a multiple looks-type model (Viemeister and Wakefield, 1991), then one would expect average pitch strength to increase proportionally to stimulus duration up to the duration of the snapshot window, after which pitch strength would

converge to a stable mean as multiple snapshots, or looks, are accumulated. This may provide explanation for the 106.07-Hz outlier thresholds observed at duration=4d in Experiment 2, as this duration may be shorter than the individual looks. An interesting future experiment would be to formally test this hypothesis.

Chapter 5

Effect of spectral resolvability on the usefulness of pitch as a cue for listening in noisy environments

I. INTRODUCTION

Listening in noisy environments can scramble the message of a signal or render it completely inaudible. For example, a tannoy announcement on a station platform can be rendered inaudible when a high-speed train passes by. In this scenario, the listener is unable to hear the speech signal at all. Alternatively, the message of a speech signal may be scrambled while conducting conversation in a noisy public place with many competing speech sources. In this scenario, the message becomes incomprehensible because the listener is unable to perceptually segregate the target signal from competing maskers, even though the target signal is well audible. Despite these difficulties, we are often able to overcome communication limitations imposed by masking, as there are a number of cues available to the auditory system that aid both detection and sound source segregation.

The first half of this chapter is concerned with how the auditory system uses monaural pitch cues to aid signal detection in noise. It is well known that interaural temporal differences (ITDs) play an important role in enhancing the detectability of signals in noise (for review, see Durlach and Colburn, 1978). When the interaural phase characteristics of the signal and masker differ, there can be a substantial masking release, even though there is no change in the signal-to-masker ratio (Hirsh, 1948). This phenomenon is referred to as the binaural masking release (BMR). While BMR has attracted much detailed investigation, the question of whether monaural pitch cues might also help to enhance signal audibility in noisy situations has been relatively neglected in the past. This is probably because under most circumstances, the effects of pitch might be confounded by other monaural cues, unrelated to pitch, such as beating in the envelope of the composite stimulus. Most tonal stimuli – for example, a pure tone or harmonic complex tone (HCT) – have periodic waveforms. Summation of periodic waveforms that have slight differences in periodicity results in relatively slow, periodic amplitude

envelope interactions. These interactions give rise to percepts such as beating and roughness. Therefore, in studies where HCT signals are presented in HCT maskers (Micheyl et al., 2006), the unmasking contributions of pitch cues are inseparable from the unmasking contributions of these envelope interaction cues. The asymmetry of masking observed between noise and HCTs (Gockel et al., 2002, Gockel et al., 2003) has also been attributed to envelope cues, in that the relatively large crest factor of the HCT envelope allows the participant to listen in the dips, whereas a noise envelope has a relatively low crest factor. Thus, in order to isolate the contribution of pitch cues to monaural masking release, envelope interactions between the signal and masker need to be minimized. To achieve this, IRN stimuli were used, because IRNs have a noise-like peak factor and non-deterministic envelopes. When uncorrelated IRNs are presented simultaneously, there is a reduction in the temporal regularity of the composite stimulus relative to the individual components. Krumbholz et al. (2003a) showed that this reduction in correlation may be used as a detection cue.

In the first part of the current study, IRN detection thresholds were measured in the presence of masking IRNs as a function of the rate difference between the components. As in previous chapters, harmonic resolvability was a parameter in these measurements. If the sensitivity of the monaural system to a reduction in serial correlation is comparable to the sensitivity of the binaural system to a reduction in interaural correlation, a sizable release from masking would be expected, even when masker and signal cannot be discriminated in terms of their pitch difference or spectral profile. As in previous chapters, listening region was also a parameter in these measurements. Krumbholz et al. (2003a) showed an effect of listening region, however the harmonic resolvability of the stimuli covaried with the listening region in which the stimuli were presented; thus, making it difficult to

disambiguate between the effects of resolvability and listening region. The aim of the first experiment was to quantify the effects of each individually.

The second half of this chapter contrasted the first half by investigating how the auditory system uses monaural pitch cues to aid simultaneous sound source segregation. For segregation to occur, components of the signal and masker must be grouped as separate auditory objects according to a common attribute. However, not all cues available to the auditory system are useful for segregation. In situations where the signal and masker are both audible, interaural timing cues are ineffective at aiding listeners to group together simultaneous spectral constituents of a composite sound source (Culling and Summerfield, 1995, Darwin and Hukin, 1999, Hukin and Darwin, 1995). In contrast, pitch is known to play an important role in simultaneous grouping (Darwin, 1981), although this is thought to be limited by peripheral harmonic resolvability (Micheyl et al., 2006). The importance of pitch cues in identification and discrimination of concurrent vowel sounds has been well established (Assmann and Summerfield, 1990, Culling and Darwin, 1993, de Cheveigné et al., 1995, Summerfield and Assmann, 1991). In the current study, listening region and resolvability were parameters under test; therefore, the use of vowel stimuli was not appropriate. This is because band-limiting vowel stimuli can disrupt the formant structure that makes the vowel identifiable. Furthermore, because of the unique spectral envelope associated with each vowel, the resolvability of individual harmonics within the stimuli changes from vowel to vowel. In order to measure the effectiveness of pitch cues for segregation as a function of rate difference between the signal and masker, we conducted a simple IRN level-discrimination experiment while simultaneously presenting a level-roved IRN masker. Contrary to common opinion, observations from this part of the study indicated that harmonic resolvability is not necessarily a prerequisite for pitch-based segregation.

II. EXPERIMENT 1: DETECTION BASED ON PITCH CUES

A. Methods

1. Stimuli

Signal and masker IRNs were generated with 16 iterations of the add-original, delay-and-add algorithm (Yost, 1996), using unity gain. The IRNs were generated in the spectral domain (for details, see Chapter 4) to avoid being limited to only using delays at integer multiples of the digital sampling period. In order to obtain the desired frequency resolution when generating the IRNs, the FFT window used for the frequency-domain filtering stages was 2^{15} samples long. This was equivalent to ~1.3 seconds of audio at a sampling rate of 25 kHz. After performing the inverse FFT, the time-domain signals were truncated to 800 ms and gated on and off with 20-ms cosine-squared ramps. Stimuli were presented using the same methods and equipment as described in previous chapters.

To assess the effect of listening region, IRNs were filtered into either a low- or high-frequency band as described in chapters 1 and 2. The low-frequency cutoff of the low-frequency band was 0.78 kHz, well within the range of phase locking. The low-frequency cutoff of the high-frequency band was 2.64 kHz. While the exact phase locking limit is unknown in humans, the fidelity of the temporal fine structure (TFS) information would have been expected to be substantially degraded in the high, compared to the low frequency listening region. The stimuli were presented in a continuous noise to mask audible distortion products below the stimulus pass-band as described in chapters 1 and 2. The repetition rates of the IRN signals were chosen according to the rule of Shackleton and Carlyon (1994), to include both unresolved (53.03 Hz and 150.00 Hz) and resolved (150.00 Hz and 424.26 Hz) IRNs in both spectral bands. For details on how these rates were derived, refer to chapters 1 and 2.

Thresholds were measured for an IRN signal in the presence of an IRN masker as a function of the IRN rate difference between the signal and masker. Rate differences between the signal and masker were quantified in cents (defined in Chapter 4). In order to best sample the features of interest in the masking patterns, thresholds were measured at 6 rate differences between signal and masker components which were log spaced between 5 and 200 cents. Masker components were always presented at a lower rate relative to each signal rate. The relationship between the signal and masker rates and their respective harmonic resolvability is represented graphically in Fig. 1.

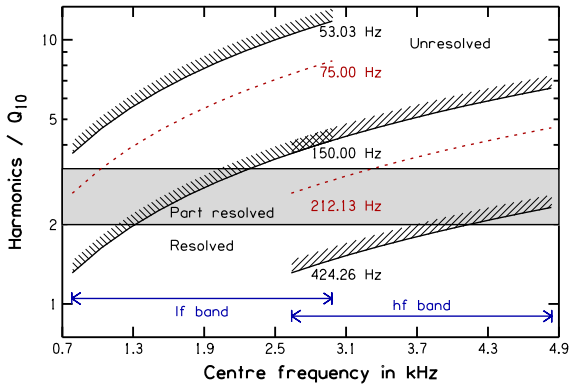

FIG. 1. Graphical representation of the parameter space used. The abscissa represents the centre frequency of the auditory filters across the listening regions in which the stimuli were presented. The ordinate represents the number of harmonics of the IRN spectra that fall into the 10-dB bandwidth of those filters. The parameter is the signal rate, where solid lines represent the signal rates used in the current study. The rates are given by the text beneath each curve. The dashed red lines represent the IRN rates corresponding to the limits of harmonic resolvability at the lower edge of each band. The shaded area shows the region between 2 to 3.25 harmonics / Q10 within which stimuli are said to be partially resolved. The wiskers protruding from the signal IRNs show the range over which the masker IRN vas varied (200 cents) relative to each signal. The asymmetry of the signal

IRNs around the limits of resolvability ensures that the higher-rate stimuli within each band always contained some resolvable components, even when the masker was 200 cents lower than the signal.

2. Procedure

The masked detection thresholds were measured using an adaptive staircase procedure. Each trial consisted of three 800-ms observation intervals, which were separated by 500-ms gaps. The masker was presented at a level of 65-dB SPL, and the signal level was varied adaptively, relative to the masker level. Two intervals contained the masker alone; the other interval contained the masker plus signal. The task was to indicate which interval contained the signal by pressing one of three response buttons. Feedback was given at the end of each trial. At the beginning of each threshold run, the signal-to-masker ratio was set to 5 dB; at this level, the signal level was well above the anticipated detection threshold for all stimulus conditions. The signal level was decreased after two consecutive correct responses and increased after each incorrect response to track the signal level that yields 70.7% correct responses (Levitt, 1971). The step size of the increments and decrements in signal level was 5 dB for the first reversal in level, 3 dB for the second reversal, and 2 dB for the rest of the eight reversals that made up each threshold run. The last six reversals of signal level were averaged to obtain a threshold estimate for each run. Three threshold runs were conducted for each participant per stimulus condition.

3. Listeners

A total of 5 listeners (4 male and 1 female, aged between 20 and 30 years) participated in the current experiments. One was the author; the others were paid for their services at an hourly rate. Participants met the criteria outlined in Chapter 1.

B. Results and interim discussion

Average detection thresholds are shown in Fig. 2. The statistical significance of the observations was tested by performing linear mixed-models analysis on the data. The analysis was performed on factors frequency band, rate difference between signal and masker IRNs, and resolvability. The dependent variable was mean threshold averaged across the three runs for each participant in each condition.

When there was no rate difference between signal and masker (rate difference = 0 cents), the only detection cue available was the loudness difference between masker alone and signal-plus-masker intervals. Mean signal to masker ratio (SMR) detection threshold averaged across all masking patterns was -3.30 dB. This corresponds to a 1.67-dB SPL difference in overall level between the signal-plus-masker and masker-alone intervals. There was a significant main effect of rate difference between signal and masker IRNs [$F(5,92)=191.798$, $p<0.001$]. Pairwise comparisons of thresholds at different rate difference showed that at a rate difference of 5 cents there was a significant reduction in thresholds and compared to when signal and masker IRNs had equal rate [$F(5,92)=191.798$, $p<0.001$]. Therefore, the addition of a pitch cue provided a release from masking for all signal rates. There was also a significant interaction between resolvability and rate difference [$F(5,92)=24.172$, $p<0.001$]. Pairwise comparisons between thresholds for resolved and unresolved stimuli were not significant when there was no rate difference between signal and masker [$F(1,92)=0.025$, $p=0.876$] , or at a rate difference of 31 cents

[F(1,92)=3.666, p=0.059]; however, thresholds were significantly different at rate differences of 5 cents [F(1,92)=29.826, p<0.001], 12 cents [F(1,92)=12.029, p<0.001], 79 cents [F(1,92)=47.249, p<0.001], and 200 cents [F(1,92)=33.717, p<0.001]. A greater masking release was observed for unresolved compared to resolved stimuli at smaller rate differences between signal and masker (5 and 12 cents); however, thresholds for unresolved stimuli did not improve significantly at rate differences greater than 12 cents: pairwise comparisons for unresolved stimuli were not significantly different between 12 and 31 cents [F(5,92)=71.054, p=0.079], 12 and 79 cents [F(5,92)=71.054, p=0.261], and 12 and 200 cents [F(5,92)=71.054, p=0.349]. Thresholds for the resolved stimuli were significantly different at all rate differences with the exception of the difference between thresholds at 31 cents and 200 cents [F(5,92)=144.916, p=0.963], suggesting that the masking pattern is non-monotonic. At rate differences of 79 and 200 cents, thresholds were lower for the resolved stimuli than for the unresolved stimuli. This may be because the listeners were able to perceptually segregate the signal IRN from the masker IRN in these conditions, rather than using perceptual changes in the sound quality alone. Section 4 of the current chapter is dedicated to quantifying the ability of listeners to segregate IRNs based on pitch cues.

There was no significant main effect of spectral band [F(1,92)=0.311, p=0.579]. However, there was a significant interaction of spectral band and resolvability [F(1,92)=19.713, p<0.001]. The unresolved IRN masking patterns followed the same basic shape as one another, but thresholds for the low-frequency band signals were ~2 dB higher overall than the high-frequency band signals. This difference is more likely to be due to the fact that 53.03 Hz is towards the lower limit of pitch perception, rather than an effect of listening region.

The differences between thresholds for resolved and unresolved stimuli are consistent with the results of Krumbholz et al. (2003a). While they did not manipulate harmonic resolvability independently of spectral region, they found sharper masking patterns in higher frequency regions where the IRNs were less resolved, but higher asymptotic thresholds. Similarly, the data from the current study shows a sharper release from masking as a function of rate difference between signal and masker for the unresolved stimuli, and higher thresholds overall for the unresolved stimuli at larger rate differences compared to the resolved stimuli. Krumbholz et al. (2003a) were able to account for their results using by measuring the differences between the time-interval histograms in response to masker-alone and signal-plus-masker stimuli generated using a modified version of the auditory image model (Patterson et al., 1995). The aim of the subsequent modelling analysis was to assess whether a similar model could account for the experimental results obtained in the current study, particularly with respect to the differences between thresholds for resolved and unresolved stimuli.

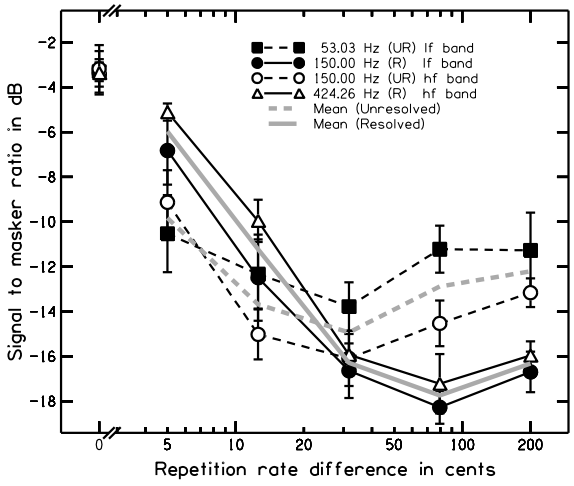

FIG. 2. Mean time interval masking patterns averaged across all listeners. Error bars show inter-listener standard error. The parameters are resolvability within each frequency band. Dashed lines show thresholds for unresolved (UR) stimuli, solid lines show

thresholds for resolved (R) stimuli, filled symbols show thresholds for stimuli presented in the low-frequency band and open symbols show thresholds for stimuli presented in the high-frequency band. SMR thresholds are plotted as a function of the rate difference between signal and masker.

C. Modelling

1. Methods and procedure

The peripheral stages and pitch-extraction mechanism of the current model were similar to that used in Krumbholz et al. (2003) and Krumbholz et al. (2001). As in previous chapters, the auditory model consisted of three cascaded stages: frequency decomposition, followed by neural transduction, followed by extraction of temporal regularity. Here, the first stage was a 49-channel gammatone filter bank with centre frequencies between 0.2 and 6 kHz, evenly distributed on the ERB scale at approximately 2 channels per ERB. In the second stage, half-wave rectification, log compression, and a 2^{nd}-order lowpass filter with a 1.2-kHz cutoff frequency was used. Finally, a channel-by-channel time-interval histogram of the neural activity in each spectral channel was produced using strobed temporal integration (STI). STI was originally designed to preserve short-term temporal asymmetry that listeners hear (Patterson and Irino, 1998). However, the output of the STI process is similar to that from the autocorrelation process used throughout the rest of the thesis and was used here for consistency with the earlier study of Krumbholz et al. (2003). It is also computationally far less expensive than an equivalent autocorrelation process. The current version of the auditory image model software (AIM) was used to generate the time interval histograms in the current study. By default, the current version of AIM applies an exponential weighting function with a half-

life of 30 ms to the time-interval histogram generated by the process of STI, reducing the level of the bins towards the lower limit of pitch.

The decision statistic was derived from the time-interval histogram, produced by averaging the time-interval histograms from each channel across frequency bands; therefore, the resulting time-interval histogram is analogous to the summary autocorrelogram (Meddis and Hewitt, 1991a, Meddis and Hewitt, 1991b). The time-interval histogram was then normalized to the value at 0 ms. Signals were generated with SMRs ranging from -24 to 0 dB in 6-dB steps and then added to the masker components. Stimuli had total durations of 2^{22} samples in order to produce very stable time-interval histograms. These were subsequently processed in frames of 2^{10} samples for computational efficiency. The first frame was omitted to remove the build-up from the impulse response of the cochlear filters. The remaining frames were then averaged to produce a single time-interval histogram. The effective-mean time-interval histogram was then produced by normalizing the mean time-interval histogram by the standard deviation of the time-interval histogram at each time-interval across frames. The standard deviation was highest in the regions between the peaks; hence, the peaks of the time-interval histogram conveyed the most information about the stimulus and were weighted higher than the surrounding background regions.

In the simulations, we used a Euclidean distance, D, to measure the differences between the effective-mean time-interval histogram for the signal plus masker and that of the masker alone. D is the square-root of the integral of the squared differences between the histograms; therefore, it includes differences at all time intervals within the histograms. For each experimental condition (each combination of spectral band, signal rate, and rate difference between signal and masker), D was calculated as a function of SMR. The conditions where signal and masker had equal rate were omitted from the

modelling process, as the listeners used loudness cues for detection, which the model was not designed to account for. Threshold was defined as the SMR at which D reached a criterion level, C, which was the main parameter in the fitting process. All of the conditions of the experiment were fitted simultaneously with a fixed value of C, and C was varied to find the value that minimized the root-mean-square (RMS) deviation between the simulated and observed thresholds.

In Chapter 3, it was shown that the autocorrelograms of IRNs were very different when generated using a dynamic compressive cochlear models, such as the pole-zero filter cascade (PZFC), as opposed to a linear model such as the GTFB. Therefore, the choice of filterbank would be expected to have a large effect on the thresholds simulated in the current study. To test this, a second set of simulated thresholds were generated using the PZFC in place of the GTFB. The instantaneous logarithmic compression was removed in this case, as compression is modelled explicitly by the PZFC. All other model parameters remained the same.

2. Modelling predictions and interim discussion

Simulated thresholds are plotted in Fig. 3. They are shown adjacent to the mean of the listeners' thresholds to aid visual comparison. The GTFB variant of the model (middle panel) was able to successfully capture the main features in the data. Thresholds decrease with increasing rate difference; resolved thresholds are slightly worse than unresolved thresholds at small rate differences but are lower at larger rate differences. This suggests that detection of an IRN signal in the presence of an IRN masker is based almost entirely on temporal pitch cues.

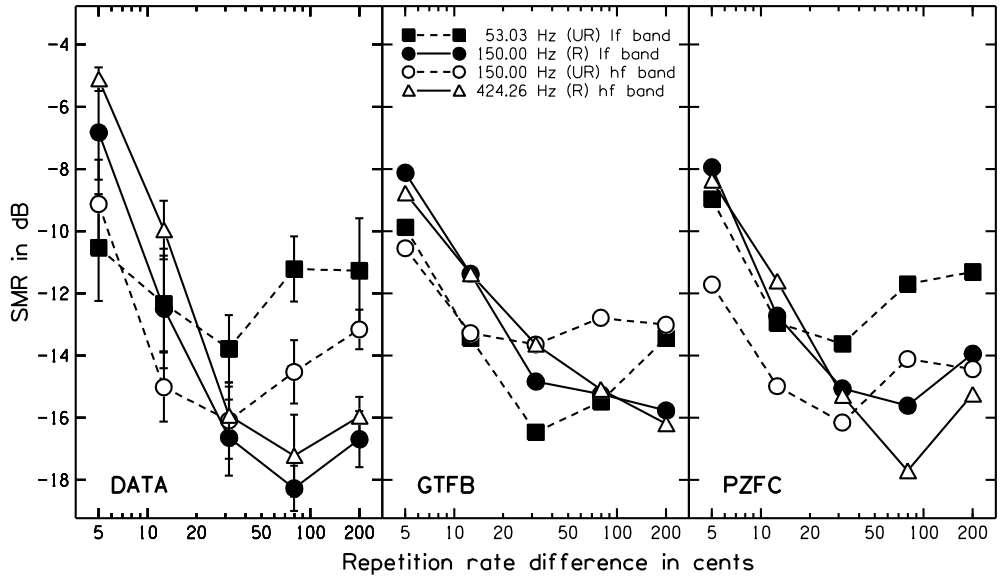

FIG. 3. Comparison of the mean data (left panel) and the simulated thresholds generated from the GTFB (central panel) and PZFC (right panel) variants of the model.

The thresholds simulated using the PZFC variant of the model matched the listener data even more closely than the GTFB variant of the model. The non-monotonic shapes of the masking patterns were well represented, as were the differences between the shapes of the resolved and unresolved patterns. Furthermore, the model also managed to predict the sensitivity difference between the unresolved thresholds observed in the data. This was likely because there were fewer peaks in the time-interval histogram of the 53.03-Hz IRN in comparison with the 150.00-Hz IRN and most of information was conveyed by the peaks.

To understand how the PZFC produced more accurate simulation results, one must compare the time-interval histograms of an IRN from both the GTFB and PZFC variants of the model. Fig. 4 shows time-interval histograms in response to a 150-Hz IRN filtered into the high frequency listening region. The background level of the GTFB time-interval histogram was much higher than the background level of the PZFC time-interval

histogram. Therefore, the information in the peaks of the PZFC time-interval histogram was weighted higher than the background. For a detailed discussion of why the background level of the PZFC time-interval histogram is lower, refer to Chapter 3.

The background levels of the time-interval histograms shown in Fig. 4. decay towards longer time intervals. This is because the overall level of the AI buffer decayed exponentially with a half-life of 30 ms, giving a lower weighting to time intervals towards the lower limit of pitch perception. This default weighting applied in AIM appears to account well for the current data without any further modification. However, in previous chapters it was shown that the auditory system is equally sensitive to modulations in pitch strength across all time intervals. Therefore, such a weighting is not appropriate in a model that can be generalized to account for both data from the current and previous chapters. An alternative theory was suggested in Chapter 1, whereby the widths of the bins that comprise the internal time-interval histogram are not equal, but greater at longer time intervals. The exponential time-interval weighting used here accounts for the data well and would be equally as effective if implemented using a logarithmic lag axis as opposed to a weighting.

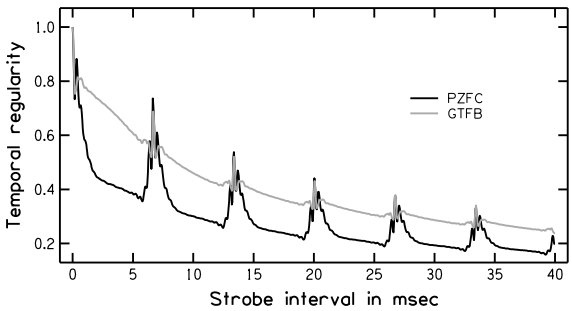

FIG. 4. Time-interval histograms of a 150-Hz IRN, where the parameter is the filter bank used. IRNs were filtered into the high spectral band to accentuate the difference between the background levels of the time-interval histograms.

III. CONTROL EXPERIMENT: SPECTRAL CONTRIBUTIONS TO THE OBSERVED MASKING RELEASE

A. Rationale

In the first experiment, a large masking release was observed when pitch cues were introduced between the signal and masker components. This was true not only for resolved, but also for unresolved stimuli, in which spectral cues would be assumed to be unavailable. In resolved stimuli, however, some of the observed masking release may have been brought about by reduction in spectral overlap. To gain a quantitative estimate of the spectral contribution to the masking release, we used the pulsation threshold technique (Houtgast, 1972), whereby an interrupted sound is perceived as being continuous, if there is sufficient energy from another sound during the interruptions. In the pulsation paradigm, masker and signal components are presented non-simultaneously, thus preventing any masking release occurring as a result of temporal interactions between the components.

The auditory continuity illusion occurs when a listener is presented with a series of alternating high- and low-level sounds. If the high-level sound (masker) has enough intensity to mask the low-level sound (signal), if they were presented simultaneously, then the signal will be perceived as continuing through the masker, despite its actual physical discontinuity. If the level difference between signal and masker does not meet this criterion, the intermittence between the signal and masker will be perceived. This phenomenon is observed readily so long as the signals are at least twice the duration of the maskers (Drake and McAdams, 1999). A pulsation threshold is defined as the highest signal level that will still give rise to the perception of continuity. This level would be expected to reflect the excitation level of the masker at the tonotopic locations of the signal.

B. Methods

1. Stimuli

The signal and masker IRNs used in this experiment were generated and filtered in the same way as those used in Experiment 1. Thresholds were measured across the same parameter space, with the exception of the conditions where the masker and signal had the same repetition rate.

Pulsation sequences were composed of temporally interleaved signal and masker IRNs with overlapping 5-ms squared-cosine cross-fade ramps to prevent audible clicks at the transitions. The sequence structure was composed of an initial 1100-ms signal to clearly identify the signal component, after which masker and signal were presented interleaved, where the masker intervals were 100 ms in duration between their -3-dB points and the signal intervals were 300 ms long. The stimulus as a whole was gated on and off with 20-ms cosine-squared ramps and had a total duration of 3000 ms.

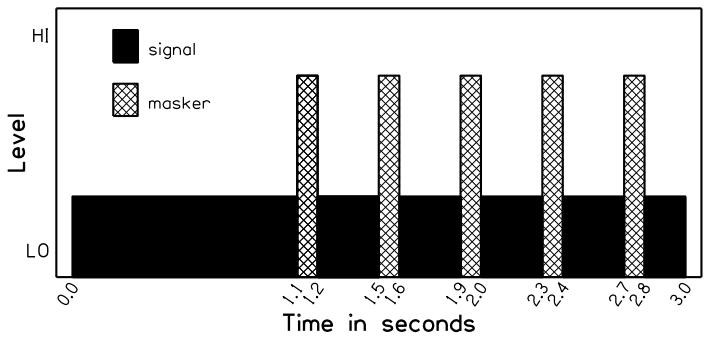

FIG. 5. Diagram depicting temporal sequence structure of the stimulus used in the current experiment. The masker depicted has a high level relative to that of the signal. Therefore, this is a condition where the signal may be perceived as continuous by the listener.

2. Procedure

Pulsation thresholds were measured using the doublet procedure (Bode and Carhart, 1973, Leek, 2001), where each threshold run consisted of 2 interleaved adaptive 2I2AFC tracks. The adaptive parameter was the level of the signal. At the beginning of each threshold run, one of the tracks had a SMR well above the anticipated continuity threshold and was tracked using a 2-down, 1-up rule. The other track began with a SMR well below the anticipated continuity threshold and was tracked using a 2-up, 1-down rule. The listeners' task was to indicate whether they perceived the signal as continuous or discontinuous by pressing one of two buttons on a response box. As this was a subjective task, no feedback was given. The step size of the changes in signal level was 5 dB in up to the first reversal in each track, 3 dB up to the second reversal, and 2 dB for the rest of the eight reversals in each track that made up one threshold run. The order of presentation of the two tracks was randomised using a weighted probability function. The final threshold estimate was the average of the last six reversals in each track, averaged across both tracks. Each participant completed three threshold runs of each stimulus condition. The listeners were the same as those who participated in Experiment 1.

C. Results and interim discussion

Thresholds averaged across listeners are presented in Fig. 6. In general, intra-listener variability was relatively small in comparison with the inter-listener variability. The statistical significance of the observations was tested by performing a linear mixed-models analysis on the data. The analysis was performed on factors rate difference, frequency band, and resolvability. The dependent variable was mean threshold averaged across the three runs for each participant per condition. This analysis revealed a significant main effect of rate difference [$F(4,76) = 8.863$, $p < 0.001$] and of resolvability

[F(1,76)=15.381, p<0.001], but in insignificant main effect of spectral band [F(1,76)=3.019, p=0.086]. The interaction between resolvability and rate difference was also significant [F(4,76)=8.386, p<0.001]. The data for the unresolved conditions was relatively flat and pairwise comparisons between thresholds at all combinations of rate differences were all insignificantly different. This confirmed that the excitation pattern of the unresolved stimuli was flat. Therefore, there would have been minimal contributions of spectral cues to the masking release observed for the unresolved stimuli in Experiment 1. Therefore the masking release observed for unresolved stimuli must have been based almost exclusively on temporal cues. The threshold at which the continuity illusion occurred for the unresolved IRNs was approximately -2 dB when averaged across all listeners. The difference between this baseline and any lower SMRs measured in this experiment can be thought of as the maximum possible spectral contribution in dB to the simultaneous masking release observed in the detection experiment. The pulsation thresholds for the resolved stimuli were dependent on the rate difference between the signal and masker and the functions were non-monotonic. Resolved and unresolved thresholds were not significantly different at rate differences of 5 cents [F(1,76)=1.826, p=0.181] and 12 cents [F(1,76)=0.288, p=0.593], but were significantly different at rate differences of 31 cents [F(1,76)=6.466, p=0.013], 79 cents [F(1,76)=35.802, p<0.001], and 200 cents [F(1,76)=4.544, p=0.036]. The lowest average threshold measured was -5.5-dB SMR at a rate difference of 79 cents. This means that the spectral contribution of resolved signals to simultaneous masking release would be assumed to be 3.5 dB at most. This is just a fraction of the ~15-dB masking-level difference observed for resolved IRNs in the simultaneous detection experiment. However, the spectral contribution to unmasking may explain the differences between thresholds for resolved and unresolved stimuli at a rate difference of 79 cents measured in Experiment 1.

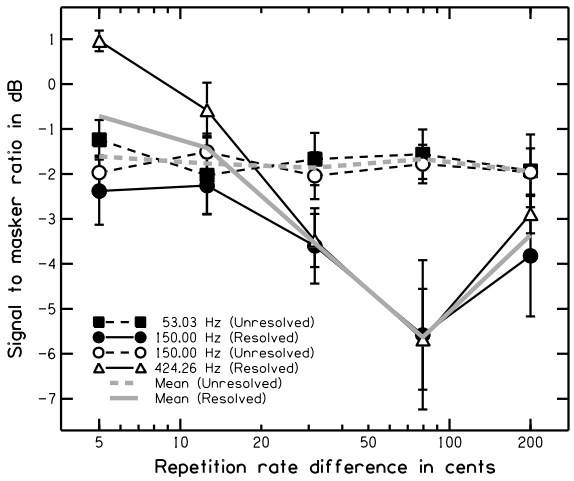

FIG.6. Mean pulsation thresholds averaged across all listeners who took part in the study. Error bars show standard error between listener means. Means of the resolved and unresolved masking patterns are also highlighted in this panel.

D. Modelling

Pulsation thresholds are thought to reflect the overlap between the internal spectra of the signal and masker components. Therefore, a simple model of the differences between the internal spectra of signal and masker components should be in agreement with the general threshold patterns observed in the data.

Signal and masker IRNs were filtered independently using a gammatone filter bank. The channel density was increased from 2 to 10 channels per ERB in order to give a better spectral resolution. Filters used were limited to those greater than half an octave below the low-frequency cutoff of the stimuli. The output of lower frequency filters would have been masked by the lowpass noise. The RMS amplitude of the signals in each channel was then calculated to produce a spectral profile of the signal and a separate spectral profile of the masker. The IRNs were 2^{22} samples in duration to obtain stable spectral representations of the stimuli. In the simulations, we used a Euclidean distance, D,

to measure the differences between the spectral profiles of each combination of signal and masker.

The simulation results are shown in Fig. 7. The values of D reflect the spectral differences between signal and masker components, and so a higher D corresponds to a lower threshold in the pulsation experiment. The values of D for the unresolvable signal and masker components are flat across the range of rate differences, as is observed in the listener data presented in Fig. 6. The general shapes of the simulated spectral differences between the resolvable signal and masker are in general agreement with the measured pulsation thresholds.

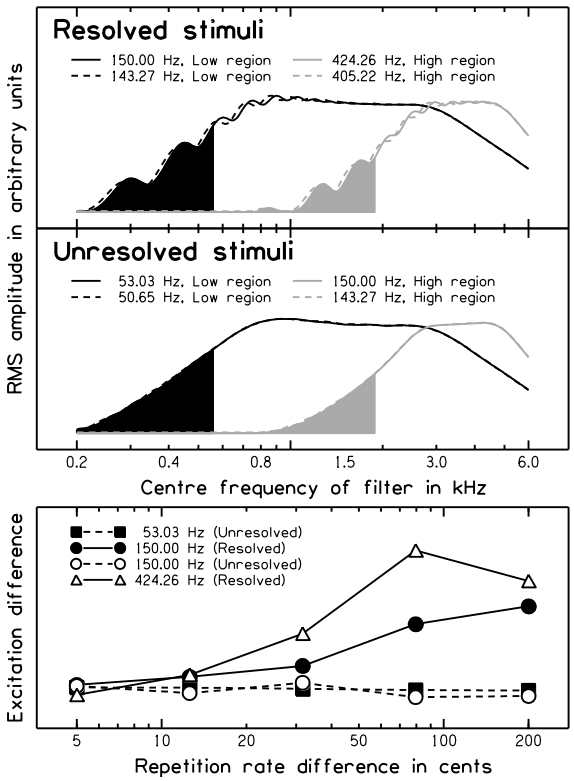

FIG.7. The upper pair of panels shows the simulated internal spectral profiles of the IRNs. The uppermost panel shows the spectral profile of IRNs containing some resolvable harmonics. Low-frequency band stimuli are shown in black, and high-frequency band stimuli are shown in grey. The solid lines represent the signal IRN, and the dashed lines

represent the masker IRN in the condition where the profiles are maximally different (79 cents rate difference between signal and masker). The bottom panel of the pair shows the same output, but for IRNs containing entirely unresolvable harmonics. Even at 79 cents there is no visible difference between unresolvable signal and masker profiles, in either frequency range. The shaded regions represent the channels less than 0.5 octaves below the stimulus cutoff. These channels were omitted from difference calculations. The separate, lowermost panel shows the RMS difference between signal and masker profiles in arbitrary units as a function of pitch difference between signal and masker IRN.

IV. EXPERIMENT 2: SEGREGATION BASED ON PITCH CUES

A. Methods

1. Stimuli

This final experiment was conducted to find how listeners can use the same pitch cues that helped them to detect the signal in Experiment 1 to perceptually segregate the signal away from the masker. The signal and masker IRNs used in this experiment were generated and filtered in the same way as those used in Experiment 1. Segregation thresholds were measured across the same range of rate differences between signal and masker. Signal and masker component were both presented around a nominal level of 60 dB SPL.

2. Procedure

Intensity discrimination thresholds were measured for an IRN signal in the presence of and IRN masker as a function of the rate difference between signal and masker IRNs. In this experiment, both signal and masker were audible, and the masker component was masking the intensity cues in the signal as opposed to reducing the detectability of the

signal itself like in Experiment 1. Level discrimination thresholds were measured using an adaptive staircase procedure. Each trial began with a 400-ms cue interval containing the signal IRN alone. This was followed by a 750-ms gap, which was followed by two 800-ms observation intervals separated by 500-ms gaps. Observation intervals contained composite signal and masker IRNs.

At the beginning of each threshold run, the signal component had an intensity difference (ΔI) of 20 dB between intervals. The higher level signal was randomly presented in one of the two observation intervals. The task was to indicate which interval contained the signal of greatest intensity, regardless of the simultaneous masker intensity, by pressing one of two response buttons. Feedback was given at the end of each trial. The ΔI was decreased after three consecutive correct responses and increased after each incorrect response to track the ΔI that yielded 79.4% correct responses. The step size for the increments and decrements in ΔI was 2 dB for the first reversal, 1.5 dB for the second reversal, and 1 dB for the rest of the eight reversals that made up each threshold run. The last six reversals of signal level were averaged to obtain a threshold estimate for each run. Participants completed three threshold runs of each experimental condition. The ΔI was limited to a maximum of 30 dB to keep overall listening levels within comfortable limits.

The masker was always presented with an intensity of either 60 +/- 7.5 dB SPL. The masker intensity difference between observation intervals was opposite to the signal intensity difference in at least one of three consecutive trials, so the listener could not achieve the 3 consecutive correct responses required for a decrease in ΔI by listening to the overall loudness of the composite stimulus in each trial. If the listeners are unable to segregate signal and masker components based on rate differences, they would base their decisions on overall intensity differences between the observation intervals and would not be able to obtain ΔI thresholds below 15 dB. If the listeners are able to segregate the

components, they would hear the signal as a separate entity, and be able to compare its loudness between observation intervals independently of the loudness of the masker. Under these conditions, listeners would be expected to obtain ΔI thresholds below 15 dB.

3. Listeners

With the exception of the author, the listeners who took part in the discrimination experiment were different to those who took part in the detection and pulsation experiments. The 4 male listeners and 1 female listener met the criteria outlined in Experiment 1.

B. Results and interim discussion

Discrimination thresholds for each individual listener and thresholds averaged across all listeners are shown in separate panels of Fig. 8. The ΔI thresholds are plotted as a function of the rate difference between signal and masker stimuli. The statistical significance of the observations was tested by performing a linear mixed-models analysis on the data. The analysis was performed on factors frequency band, rate difference between signal and masker IRNs, and resolvability. The dependent variable was mean threshold averaged across the three runs for each participant in each condition.

When there was no rate difference between signal and masker components mean thresholds were all at 15 dB as would be expected when listeners were basing decisions on overall loudness. There was a significant main effect of rate difference between signal and masker components [$F(5,92)=51.508$, $p < 0.001$]. However, in contrast to the detection results of Experiment 1, pairwise comparisons between thresholds at rate differences of 0 and 5 cents were not significant [$F(5,92)=51.508$, $p=.573$]. This suggests that listeners can

benefit from small rate differences when detecting a signal, but are unable to exploit the same pitch cue for simultaneous segregation.

The analysis revealed a significant main effect of resolvability on thresholds [$F(1,92)=112.196$, $p<0.001$], suggesting that resolved and unresolved thresholds are different and no significant main effect of spectral band [$F(1,92)=0.015$, $p=0.904$]. There was also a significant interaction between rate difference and resolvability [$F(5,92)=18.172$, $p<0.001$] and no significant interaction between rate difference and band [$F(5,92)=1.763$, $p=0.128$]. Listeners began to benefit from rate-difference cues at ~12 cents when the stimuli were resolved, as shown by the significant pairwise comparison between thresholds for resolved stimuli between 5 and 12 cents [$F(5,92)=51.263$, $p=0.002$]. Thresholds for resolvable IRNs decreased rapidly as the rate difference was increased from 12 to 79 cents. Larger rate differences provided little additional benefit to segregation performance as the difference between thresholds at 79 and 200 cents was not significant [$F(5,92)=51.263$, $p=0.776$]. This finding is in agreement with results from concurrent vowel studies (Assmann and Summerfield, 1990, Scheffers, 1983, Zwicker, 1984) which have generally shown that vowel identification performance increases rapidly with rate differences between signal and masker vowels up to intervals of about 1 semitone (100 cents). Larger rate differences provide little additional benefit and performance improves very little, if at all, compared to thresholds at 1 semitone. Until now, pitch based segregation has only been considered possible in stimuli containing resolved harmonics (for review, see Micheyl et al., 2006).

The most striking result of the current study is that harmonic resolvability is not a prerequisite for segregation in this simple task. Unlike the resolved stimuli, pairwise comparisons between thresholds for unresolved stimuli at rate differences less than 200 cents were not significantly different from one another. However, thresholds were

significantly different between 200 cents, and thresholds at all other rate differences [$F(5,92)=18.417$, $p<0.001$]. Taken together, this suggests that on average, listeners require larger rate differences between the signal and masker components to perform segregation using unresolved stimuli (~200 cents) than using resolved stimuli (~12 cents). Importantly, they were still able to utilise pitch cues in unresolved stimuli to perform segregation. Thresholds were still significantly lower for the resolved stimuli compared to the unresolved stimuli at a rate difference of 200 cents [$F(1,92)=11.415$, $p=0.001$]. Thresholds for each individual listener (smaller panels of Fig. 8.), show that while the asymptotic thresholds of each listener were similar, there is high inter-listener variability in the rate differences required for each individual to perform segregation.

Under certain conditions, people can attend to one sound within an auditory scene as if it were presented by itself. Bregman (1994) has described this as the "transparency of sound". This phenomenon was apparent in the current data, as some listeners (participants 2 and 4 in Fig. 8.) were able to discriminate level differences of just over 1 dB at the larger rate differences used in this study. This is similar to the level discrimination threshold that would be expected if the signal were presented on its own.

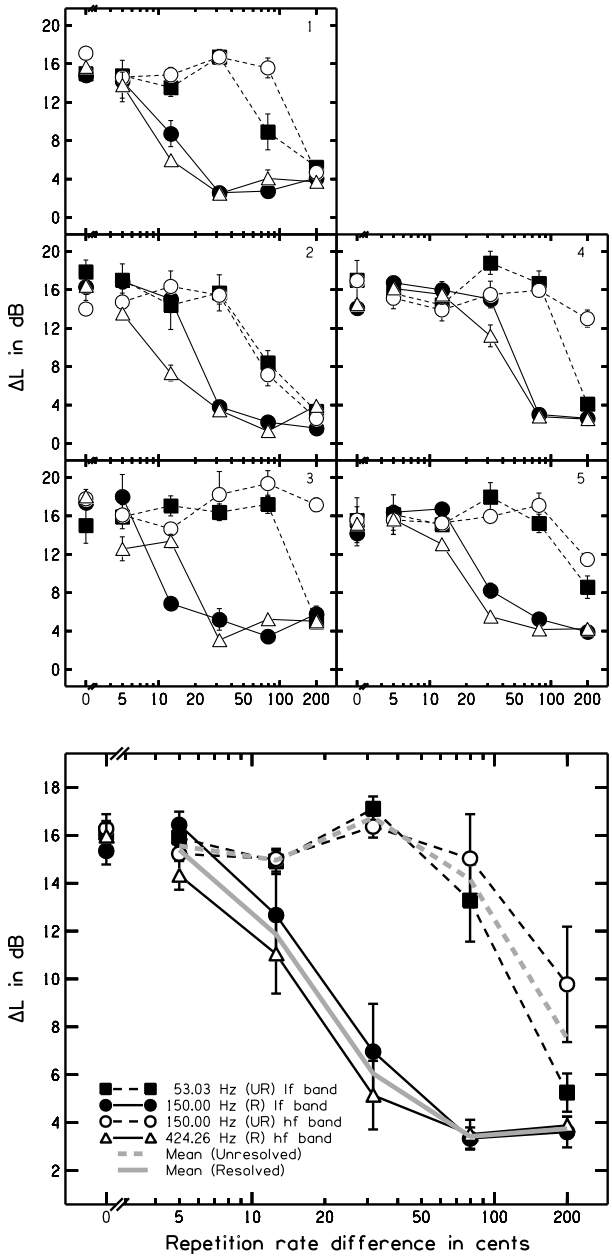

FIG.8. The smaller panels show mean thresholds for each listener as a function of rate difference between the signal and masker. Error bars represent intra-listener standard error. The larger panel at the bottom of the figure shows threshold patterns averaged across all listeners, where error bars show inter-listener standard error. Means of the resolved and unresolved masking patterns are shown by the gray lines.

C. Modelling

Models based on channel assignment have been designed to account for results of concurrent vowel experiments, in which the identification scores of the individual vowels increase when a pitch cue is introduced between the vowels (Assmann and Summerfield, 1990, Scheffers, 1983, Zwicker, 1984). The channel assignment model of Meddis and Hewitt (1992) consists of a peripheral simulation, after which the signal within each channel is subjected to an autocorrelation. Channels are then assigned into groups according to the temporal interval at which there is greatest correlation. This gives separate multi-channel representations of both signal and masker components. By definition, unresolved stimuli contain many components per channel. Therefore, a channel assignment model would be expected to have more difficulty in separating unresolved stimuli. The aim of the current analysis was to test this.

Signal and masker IRNs that were used in the current study were combined at equal RMS levels and then filtered using a gammatone filter bank with 10 channels per ERB to give a high spectral resolution. Each channel was then half-wave rectified, lowpass filtered, and compressed using logarithmic compression to simulate neural transduction. Time interval histograms were generated in response to the signal within each channel using STI. However, the NAP within each channel was retained. Channels with greater activation at the time interval corresponding to signal rate were assigned to the signal, and those with greater activation at the time interval corresponding to the masker rate were assigned to the masker. Temporal information at non-integer multiples of the sample rate was included by linearly interpolating between neighbouring sampling points. A NAP waveform was then produced for both the signal and masker by summing the composite NAP channels assigned to the signal and masker separately.

To quantify the segregation performance of the model, the signal and masker groups each had to be matched to some kind of template. Unlike the vowel stimuli used by Meddis and Hewitt (1992), the plain IRNs used in the current study did not carry unique, identifiable spectral profiles that could be matched to a template. Fujiki et al. (2002) developed a novel tagging technique, designed to extract the ratio of inputs presented to each ear from neuromagnetic responses in each hemisphere of the human auditory cortex. Their method involved sinusoidal amplitude modulation of the signals presented to each ear at slightly different modulation rates. The contribution of information from ipsilateral and contralateral inputs could then be extracted separately for each hemisphere of the cortex by taking the FFT of the neuromagnetic waveforms and calculating the ratio of the magnitude of the FFT components at the modulation frequencies. This technique was applied to the stimuli in the current model by modulating the signal at 22.89 Hz and the masker at 24.41 Hz. The modulation frequencies chosen corresponded to the centre frequencies of non-adjacent bins in a 2^{15} point FFT. The time-interval histograms were limited to 40 ms; therefore, they were not affected by the sub-25-Hz modulations. Both signal and masker were modulated at full modulation depth before they were summed and presented to the model.

To extract the tags from the segregated signal and masker NAPs, FFT spectra were calculated in rectangular windows of 2^{15} samples for each of the signal and masker NAPs. The FFT window was moved in steps of 2^{14} samples between calculations. The total stimulus duration was 2^{22} samples; therefore, 2^{8} spectra were averaged to improve the signal-to-noise ratio. Segregation performance was defined as the ratio of the magnitudes of the FFT spectrum at the tag frequencies.

The segregation performance of the model is presented in the upper panel of Fig. 9. As expected, the model was completely unable to separate unresolved signal and masker

IRNs, even at the largest rate differences between the components. Therefore, there were approximately equal amounts of signal and masker components within each group of separated channels at all rate differences. The model began to segregate the resolved 424.26-Hz signal from the masker by a rate difference of 31.62 cents. The model began to segregate the resolved 150.00-Hz signal from by a rate difference of 79.53 cents. Segregation performance improved for both resolved stimuli at larger rate differences between the components. Overall, the model was better able to segregate the 424.26-Hz stimulus compared to the resolved 150.00-Hz stimulus.

The model performance was quite different to the listener performance measured experimentally. Firstly, harmonic resolvability was not a prerequisite for segregation in the listener data. Secondly, segregation performance was statistically similar for both 150.00- and 424.26-Hz resolved IRNs, and segregation began to occur at rate differences as small as 12.57 cents between signal and masker in the listener data. Taken together, this suggests that listeners were performing perceptual segregation of signal and masker IRNs using within-channel temporal information, not only for the unresolved, but also for the resolved stimuli, perhaps in addition to a channel assignment process.

The lower panel of Fig. 9 illustrates how segregation may be performed based on the information present in the time-interval histogram of an unresolved composite stimulus. At small rate differences, the signal and masker peaks become fused when averaged across frequency channels. The fused peaks provide useful information for a signal detection model, as the fused peak height of a composite stimulus is reduced relative that of a single IRN (Krumbholz et al., 2003a). However, the fused peak does not provide any information about the individual signal and masker components. At larger rate differences, the masker and signal peaks begin to separate, becoming two distinct peaks in the time-interval histogram. The height of each of these peaks conveys information about

the relative level of each component. Therefore, the time-interval histogram contains enough information to perform the level discrimination task conducted in the current study, even for unresolved stimuli.

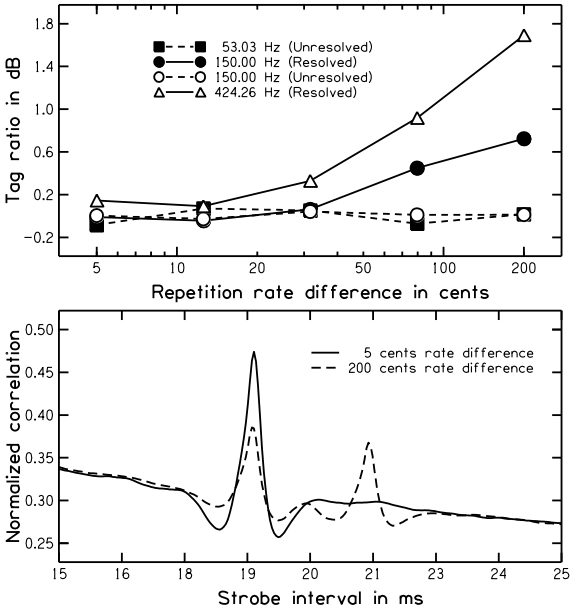

FIG.9. The upper panel shows the level ratio of the tags recovered from the signal and masker components after being separated by the model. This performance metric is plotted as a function of rate difference between signal and masker. The lower panel shows the SAI of the 53-Hz unresolved signal IRN merged with an equal level masker IRN at rate differences of 5 cents (solid line) and 200 cents (dashed line). Only lags around the first peak are shown. The peak of the signal IRN is clear in both functions at ~18 ms. The peak of the masker component is merged with the signal component when the rate difference is 5 cents. The peak of the masker component is clearly separated on the lag axis from the peak of the signal component when the rate difference is 200 cents.

V. DISCUSSION

The first experiment in the current chapter expanded upon the experimental findings of Krumbholz et al. (2003a). This was achieved by systematically assessing the effects of the harmonic resolvability on detection thresholds for IRN signals in the presence of IRN maskers as a function of the rate difference between components. Not only was a release from masking observed for the resolved stimuli, but a sizable release from masking was also observed for the unresolved stimuli. This indicated that detection was primarily based on the temporal information rather than the differences between the spectral profiles of the signal and masker components. As expected, results from the control experiment showed that the unresolved stimuli provided no spectral cues that could be used to provide a masking release. The spectral contribution from the resolved stimuli could only account for a small proportion (~3.5 dB) of the masking release observed in the simultaneous detection experiment (~15 dB), thus providing further evidence that detection thresholds were primarily based on the temporal information in the stimuli.

The masking-level differences observed in Experiment 1 were similar in magnitude to the binaural masking-level difference, suggesting that monaural pitch-based unmasking probably involves a similar processing mechanism to that responsible for providing binaural unmasking. Krumbholz et al. (2003) were able to successfully account for their masking data based on the differences between the time-interval histograms generated in response to the masker alone and signal-plus-masker. Their model was very similar to the GTFB model presented in Experiment 1 of the current study. Modifications to the model were suggested by Krumbholz et al. (2003) in order to better account for their experimental data. These modifications included adjusting the weighting applied to the time-interval histograms and also limiting of the number of peaks in the time-interval

histogram that contributed to the difference calculation. The predictions of the detection thresholds measured in the current study were acceptable when using the GTFB variant of the model, but were astoundingly accurate when using the PZFC variant of the model. No peak-order limits were imposed, suggesting that the internal decision mechanism optimally differentiates between time-interval histograms. An exponential weighting was applied in the default version of the STI model used in the current study to reduce information in the time-interval histogram towards the lower limit of pitch. However, a logarithmic lag axis (discussed in Chapter 1) would be expected to be equally effective.

An important feature of the model was that it was able to account for the main differences between the observed masking patterns for resolved and unresolved stimuli. Marked differences between pitch-discrimination performance for resolved and unresolved tonal stimuli (Carlyon, 1996b, Carlyon and Shackleton, 1994, Houtsma and Smurzynski, 1990) have been used as justification for the coexistence of spectral and temporal pitch-extraction mechanisms. In the aforementioned studies, the superior performance of listeners in conditions where stimuli contained resolved harmonics is strongly suggestive of a spectral mechanism that breaks down for unresolved stimuli. The model presented in the current study is able to explain the main resolvability-dependent differences between thresholds using a temporal pitch-extraction mechanism alone, thus providing no evidence for the involvement of a spectral pitch-extraction mechanism in the observed masking release. However, comparing the modelled thresholds to the listener data (Fig. 3), at rate differences of 79 and 200 cents between the signal and masker, the model did not predict quite as large a masking release for the resolved thresholds as was observed in the listener data. The model presented did not use information based on the differences between the spectral profiles of masker alone and signal-plus-masker, and the control experiment showed the contribution of spectral cues to be small relative to the

overall masking release. However, inclusion of a complementary spectral-profile comparison mechanism in the model may have further reduced the deviation between simulated and measured thresholds.

The last experiment presented in the current study investigated pitch-based simultaneous sound-source segregation in the presence and absence of spectral cues. Primitive grouping based on harmonicity cues has been investigated in a group of studies where listeners had to identify mistuned components in a harmonic complex (for review, see Darwin and Carlyon, 1995). The results of these studies generally suggested that listeners are only able to perceive mistuned partials as a separate auditory object when the harmonic number is low. At higher harmonic numbers, the mistuned partial was increasingly less resolved from its neighbouring partials, and listeners reported that they had used roughness as the detection cue. They still perceived a single auditory object, even when the partial was mistuned by an amount several times that necessary for the mistuning of the partial to be detected.

Segregation of concurrent HCTs has also been studied using fundamental-frequency (F0) discrimination paradigms (Carlyon, 1996a, Carlyon, 1997). The earlier of these studies measured listeners' performance in a sequential F0 discrimination task between consecutive HCTs. Performance was compared when the signal HCT was presented in the presence or absence of a simultaneous masker HCT. The masker was filtered into the same spectral band as the signal and had the same F0 in the two observation intervals. Signals were presented with F0s that were either higher or lower than that of the masker. The rate differences used between signal and masker were logarithmically spaced between 8.6 and 256.9 cents. Note that this range of rate differences was very similar to that used in the current experiment. In the condition where the signal and masker both consisted primarily of resolved harmonics, performance was

only moderately affected by the masker. In the condition where the signal and masker only contained unresolved harmonics, listeners reported hearing a single "crackly" sound but were still able to perform the task. Carlyon concluded that segregation of unresolved signals was probably based on the discrimination of global changes in the pitch evoked by the signal-plus-masker mixture, rather than the pitch of the signal alone. Specifically, the rate of envelope peaks of the combined masker and signal increased with increases in the signal F0. The later study by Carlyon (1997) showed that when the envelope rate cue was neutralized by using pseudorandom pulse trains instead of HCTs, performance was reduced to chance. This suggested that listeners could not accurately extract the pitch value of a signal in the presence of a tonal masker when stimuli were unresolved. In the current study, periodic envelope interaction cues were not available due to the stochastic nature of the IRN stimuli used. However, listeners were still able to perform the task when the stimuli were unresolved. The evidence suggests that listeners are able to perform pitch-based segregation in the absence of spectral cues, so long as the rate difference between signal and masker is sufficiently large. Like in concurrent vowel experiments, the current study measured segregation performance based on a feature (level) of the stimulus that was independent of the cue used for segregation (pitch). In Carlyon's experiments (1996, 1997) segregation performance was based on the discrimination of the same cue used for segregation. This fundamental difference between the tasks may explain the disagreement between experimental findings.

Segregation thresholds for resolved and unresolved stimuli behave differently as a function of the relative rate difference between signal and masker components. However, thresholds for both of the150.00- and 424.26-Hz resolved IRN signals scale together when plotted as a function of the relative rate difference between signal and masker components. Similarly, thresholds for the 50.03- and 150.00-Hz unresolved signals scale well according

to the relative rate difference between components. Generally, current autocorrelation-based models of pitch perception use a linear time axis. Therefore, a model based on the separation of signal and masker peaks in the time-interval histograms of the stimuli would predict that thresholds would scale according to the linear rate difference between components. When modelling data from the detection experiment, a lower weighting was applied to the longer time intervals in the time-interval histogram. In a segregation model, this weighting would predict a lower sensitivity to level differences in lower-pitched stimuli. However, equal sensitivity was observed. These arguments add further weight to the idea that the time intervals in the time-interval histogram should be logarithmically spaced.

GENERAL CONCLUSIONS

The current thesis comprised five studies investigating the role of temporal and spectral harmonicity cues in pitch extraction under important stimulus conditions related to listening in multi-talker environments. Within each study, the availability of spectral cues was varied, providing insight into how pitch is extracted under each condition.

The temporal resolution of both the binaural system in response to changes in binaural parameters (Akeroyd and Summerfield, 1999) and the monaural system in response to changes in intensity (for review, see Eddins and Green, 1995, Viemeister and Plack, 1993) have been thoroughly investigated. In contrast, the temporal resolution of the monaural pitch-extraction mechanism has received very little attention (Wiegrebe, 2001). Until now, no studies have assessed the role of harmonic resolvability on the temporal resolution of pitch extraction. In Chapter 1, a novel stimulus was presented, allowing the standardized measures of temporal resolution often used in the binaural and intensity domains to be measured in the pitch domain. Results suggested that the time constants of the integration window presumed responsible for limiting the resolution of monaural pitch extraction scaled according to the rate of the stimulus. The results also suggested that the pitch-related time constants were much longer than those associated with monaural intensity resolution, and thus have more in common with the time constants measured in binaural processing.

In Chapter 2, the temporal resolution of pitch extraction was measured in a higher-frequency region in which the fidelity of the TFS available to the brain was assumed to be severely degraded relative to that in Chapter 1. Much larger time constants were required to model the high- compared to the low-frequency region data, suggesting that the pitch-related time constants not only scale with pitch value, but also with frequency region. The data from the two frequency regions measured was not enough to determine the time

constants in a band of arbitrary centre frequency. Therefore, a second experiment was conducted in which gap-detection thresholds were measured over a range of centre frequencies. Results from the second study revealed that the relationship between time constants and centre frequency resembled an inverted lowpass filter function with a cutoff of approximately 1 kHz. This coincides with the frequency at which phase locking is thought to break down in humans. Therefore, the increase in time constants may reflect the system compensating for the reduction in high fidelity TFS towards higher frequencies.

Frequency region is known to have an effect on the subjective pitch strength. Given that Chapter 2 showed how the time constants of pitch extraction depend on the frequency region, the experiments in Chapter 3 were conducted to see whether the time constants of pitch extraction are dependent on the pitch strength of the stimuli when the overall pitch strength is varied by changing n, rather than the frequency region. Results from Chapter 3 suggested that time constants do not vary according to the pitch strength, and that the results could be modelled by a fixed time constant, so long as sensitivity differences between stimuli with different n were accounted for using the expansive function suggested by Yost (1996). The second part of the chapter considered the implications of cochlear compression on how expansion should be modelled in a neural model of pitch strength.

No effects of harmonic resolvability were observed in any of the first three chapters measuring the temporal resolution of pitch extraction. This suggests that the pitch-extraction mechanism responsible for limiting temporal resolution is either based entirely on a temporal mechanism, or spectral and temporal mechanisms that feed into a common integrator, or that the integrators associated with spectral and temporal mechanisms are functionally identical.

The effect of stimulus duration on pitch-value discrimination thresholds has been used to quantify the duration over which pitch information can be integrated. The assumption has been that discrimination thresholds reach asymptote at the stimulus duration corresponding to the length of the pitch integration window. However, the time constants derived from the resolution data measured in the high-frequency band used in Chapter 2 were much longer than those measured in an earlier integration task (White and Plack, 2003) in which stimuli were presented in a similar band. This paradoxical result motivated the experiments presented in Chapter 4. The effect of stimulus duration on thresholds was compared in both pitch-strength and pitch-value discrimination tasks. Thresholds measured in the pitch-value discrimination task reached asymptote by approximately 8 stimulus cycles, which was in close agreement with results from similar previous studies (Krumbholz et al., 2003b, White and Plack, 2003); however, the pitch-strength discrimination task showed performance was only limited by the stimulus duration. Taken together, the results from the different tasks suggested that the duration at which thresholds reach asymptote may not truly represent the integration capacity of the system. Modelling suggested that the relationship between discrimination thresholds and stimulus duration may only reflect the variance within the internal estimate pitch value or pitch strength.

The data presented in Chapters 1 – 4 showed no effects of harmonic resolvability. However, this may have been because the stimuli were presented in quiet backgrounds. Pitch is well known to be one of the most important cues for simultaneous grouping of concurrent sounds (Darwin, 1981), and the availability of spectral cues is thought to be a prerequisite for segregation of simultaneous sound sources to occur. Pitch cues have also been shown to aid detection of a tonal signal in the presence of a tonal masker (Krumbholz et al., 2003a). In Chapter 5, the masking release obtained from pitch cues was measured.

The optimum masking release was shown to be approximately 15 dB, suggesting that the processing mechanism responsible for providing the pitch-based masking release is similar to that responsible for binaural unmasking. Most of the observed masking release could be accounted for using a temporal model of pitch, and the subtle differences between modelled and measured thresholds could be explained by the spectral contributions to unmasking measured in the control experiment. This strongly indicates that pitch-based unmasking is mostly based on a temporal pitch-extraction mechanism. In contrast to the first part of the study, the second part measured how pitch cues aid simultaneous grouping in the presence and absence of spectral cues. Contrary to common assumption, data from this part of the study revealed that harmonic resolvability was not a prerequisite for segregation to occur, suggesting the need for a temporal model of segregation based on the separation of peaks in the time interval histogram to complement current models based on spectral channel assignment (Meddis and Hewitt, 1992).

The current work has provided new insights on how pitch is extracted by the auditory system. Importantly, almost all of the data presented could be accounted for by temporal models of pitch extraction, stressing the importance of the availability of temporal pitch information to the brain, even in high-frequency regions where the fidelity of temporal information is known to be degraded. The current work has also highlighted a number of parallels between the processing of pitch and binaural temporal processing. Due to the limited spectral resolution available in cochlear implants, it is particularly important to encode temporal information as effectively as possible, and the results contained in this thesis may have implications for such work.

REFERENCES

AKEROYD, M. A. & SUMMERFIELD, A. Q. 1999. A binaural analog of gap detection. The Journal of the Acoustical Society of America, 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. The Journal of the Acoustical Society of America, 49**,** 1131-1139.

ASSMANN, P. F. & SUMMERFIELD, Q. 1989. Modeling the Perception of Concurrent Vowels - Vowels with the Same Fundamental-Frequency. Journal of the Acoustical Society of America, 85**,** 327-338.

ASSMANN, P. F. & SUMMERFIELD, Q. 1990. Modeling the Perception of Concurrent Vowels - Vowels with Different Fundamental Frequencies. Journal of the Acoustical Society of America, 88**,** 680-697.

BALAGUER-BALLESTER, E., CLARK, N. R., COATH, M., KRUMBHOLZ, K. & DENHAM, S. L. 2009. Understanding Pitch Perception as a Hierarchical Process with Top-Down Modulation. PLoS Computational Biology, 5.

BODE, D. L. & CARHART, R. 1973. Measurement of articulation functions using adaptive test procedures. IEEE Transactions on Audio and Electroacoustics, AU21**,** 196-201.

BREGMAN, A. 1994. Auditory scene analysis: The perceptual organization of sound, (MIT Press, Cambridge, MA).

BUUNEN, T. J. F. & VAN VALKENBURG, D. A. 1979. Auditory detection of a single gap in noise. The Journal of the Acoustical Society of America, 65**,** 534-537.

CARLYON, R. P. 1996a. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. The Journal of the Acoustical Society of America, 99**,** 517-524.

CARLYON, R. P. 1996b. Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics. The Journal of the Acoustical Society of America, 99**,** 525-533.

CARLYON, R. P. 1997. The effects of two temporal cues on pitch judgments. The Journal of the Acoustical Society of America, 102**,** 1097-1105.

CARLYON, R. P. 1998. Comments on "A unitary model of pitch perception" [J. Acoust. Soc. Am. [bold 102], 1811--1820 (1997)]. The Journal of the Acoustical Society of America, 104**,** 1118-1121.

CARLYON, R. P. & SHACKLETON, T. M. 1994. Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms? The Journal of the Acoustical Society of America, 95**,** 3541-3554.

CULLING, J. F. & DARWIN, C. J. 1993. Perceptual separation of simultaneous vowels - within and across-formant grouping by F0. Journal of the Acoustical Society of America, 93**,** 3454-3467.

CULLING, J. F. & SUMMERFIELD, Q. 1995. Perceptual Separation of Concurrent Speech Sounds - Absence of across-Frequency Grouping by Common Interaural Delay. Journal of the Acoustical Society of America, 98**,** 785-797.

CULLING, J. F. & SUMMERFIELD, Q. 1998. Measurements of the binaural temporal window using a detection task. Journal of the Acoustical Society of America. 103, 3540–3553.

DARWIN, C. J. 1981. Perceptual grouping of speech components differing in fundamental-frequency and onset-time. Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 33**,** 185-207.

DARWIN, C. J. & CARLYON, R. P. 1995. Auditory grouping, Academic Press, Inc.; Academic Press Ltd.

DARWIN, C. J. & HUKIN, R. W. 1999. Auditory objects of attention: The role of interaural time differences. Journal of Experimental Psychology-Human Perception and Performance, 25**,** 617-629.

DE CHEVEIGNÉ, A. 1998. Cancellation model of pitch perception. Journal of the Acoustical Society of America, 103**,** 1261-1271.

DE CHEVEIGNÉ, A., MCADAMS, S., LAROCHE, J. & ROSENBERG, M. 1995. Identification of concurrent harmonic and inharmonic vowels - a test of the theory of harmonic cancellation and enhancement. Journal of the Acoustical Society of America, 97**,** 3736-3748.

DENHAM, S. 2005. Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model. Biosystems, 79**,** 199-206.

DRAKE, C. & MCADAMS, S. 1999. The auditory continuity phenomenon: Role of temporal sequence structure. Journal of the Acoustical Society of America, 106**,** 3529-3538.

DUQUESNOY, A. J. 1983. Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. The Journal of the Acoustical Society of America, 74**,** 739-743.

DURLACH, N. & COLBURN, H. 1978. Binaural phenomena. In: E. Carterette, and M. Friedman (Eds.), Handbook of Perception. Volume IV: Hearing. (Academic Press, New York), 365–466.

EDDINS, D. & GREEN, D. 1995. Temporal integration and temporal resolution. In: B.C.J. Moore, (Ed.), Hearing, (Academic Press, San Diego), 207–242.

FASTI, H. (1988) Pitch and pitch strength of peaked ripple noise. In: H. Duifhuis, J.W. Horst and H.P. Wit (Eds.), Basic Issues in Hearing. (Academic Press, San Diego), 370–379.

FITZGIBBONS, P. J. & WIGHTMAN, F. L. 1982. Gap detection in normal and hearing-impaired listeners. The Journal of the Acoustical Society of America, 72**,** 761-765.

FLORENTINE, M., BUUS, S. & POULSEN, T. 1993. The growth of loudness of brief sounds. The Journal of the Acoustical Society of America, 93**,** 2367-2367.

FORREST, T. G. & GREEN, D. M. 1987. Detection of partially filled gaps in noise and the temporal modulation transfer function. The Journal of the Acoustical Society of America, 82**,** 1933-1943.

FUJIKI, N., JOUSMAKI, V. & HARI, R. 2002. Neuromagnetic responses to frequency-tagged sounds: a new method to follow inputs from each ear to the human auditory cortex during binaural hearing. Journal of Neuroscience**,** 205.

GARNER, W. R. & WERTHEIMER, M. 1951. Some Effects of Interaural Phase Differences on the Perception of Pure Tones. The Journal of the Acoustical Society of America, 23**,** 664-667.

GLASBERG, B. R. & MOORE, B. C. J. 1986. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. The Journal of the Acoustical Society of America, 79**,** 1020-1033.

GLASBERG, B. R. & MOORE, B. C. J. 1990. Derivation of Auditory Filter Shapes from Notched-Noise Data. Hearing Research, 47**,** 103-138.

GLASBERG, B. R. & MOORE, B. C. J. 2000. Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. The Journal of the Acoustical Society of America, 108**,** 2318-2328.

GOCKEL, H., MOORE, B. C. J. & PATTERSON, R. D. 2002. Asymmetry of masking between complex tones and noise: The role of temporal structure and peripheral compression. Journal of the Acoustical Society of America, 111**,** 2759-2770.

GOCKEL, H., MOORE, B. C. J. & PATTERSON, R. D. 2003. Asymmetry of masking between complex tones and noise: Partial loudness. Journal of the Acoustical Society of America, 114**,** 349-360.

GOLDSTEIN, J. L. 1973. Optimum Processor Theory for Central Formation of Pitch of Complex Tones. Journal of the Acoustical Society of America, 54**,** 1496-1516.

GRANTHAM, D. W. 1982. Detectability of Time-Varying Inter-Aural Correlation in Narrow-Band Noise Stimuli. Journal of the Acoustical Society of America, 72**,** 1178-1184.

GRANTHAM, D. W. 1984. Discrimination of dynamic interaural intensity differences. The Journal of the Acoustical Society of America, 76**,** 71-76.

GRANTHAM, D. W. & WIGHTMAN, F. L. 1978. Detectability of varying interaural temporal differences. The Journal of the Acoustical Society of America, 63**,** 511-523.

GRANTHAM, D. W. & WIGHTMAN, F. L. 1979. Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation The Journal of the Acoustical Society of America, 65, 1509–1517.

GREEN, D. M. & SWETS, J. A. 1966. Signal detection theory and psychophysics, (Wiley, New York).

HIRSH, I. J. 1948. The Influence of Interaural Phase on Interaural Summation and Inhibition. The Journal of the Acoustical Society of America, 20**,** 536-544.

HOUTGAST, T. 1972. Psychophysical evidence for lateral inhibition in hearing. Journal of the Acoustical Society of America, 51**,** 1885-1894.

HOUTSMA, A. J. M. & SMURZYNSKI, J. 1990. Pitch identification and discrimination for complex tones with many harmonics. The Journal of the Acoustical Society of America, 87**,** 304-310.

HUKIN, R. W. & DARWIN, C. J. 1995. Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. The Journal of the Acoustical Society of America, 98**,** 1380-1387.

KOLLMEIER, B., GILKEY, R. H. & SIEBEN, U. K. 1988. Adaptive staircase techniques in psychoacoustics - a comparison of human data and a mathematical-model. Journal of the Acoustical Society of America, 83**,** 1852-1862.

KOLLMEIER, B. & GILKEY, R. H. 1990. Binaural forward and backward masking: Evidence for sluggishness in binaural detection. Journal of the Acoustical Society of America, 87, 1709–1719.

KRUMBHOLZ, K., PATTERSON, R. D. & NOBBE, A. 2001. Asymmetry of masking between noise and iterated rippled noise: Evidence for time-interval processing in the auditory system. Journal of the Acoustical Society of America, 110**,** 2096-2107.

KRUMBHOLZ, K., PATTERSON, R. D., NOBBE, A. & FASTL, H. 2003a. Microsecond temporal resolution in monaural hearing without spectral cues? Journal of the Acoustical Society of America, 113**,** 2790-2800.

KRUMBHOLZ, K., PATTERSON, R. D., SEITHER-PREISLER, A., LAMMERTMANN, C. & LUTKENHONER, B. 2003b. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cerebral Cortex, 13**,** 765-772.

LEEK, M. R. 2001. Adaptive procedures in psychophysical research. Perception and Psychophysics, 63**,** 1279-1292.

LEVITT, H. 1971. Transformed Up-Down Methods In Psychoacoustics. Journal of the Acoustical Society of America, 49**,** 467-477.

LICKLIDER, J. C. R. 1951. A Duplex Theory Of Pitch Perception. Experientia, 7**,** 128 - 134.

LORENZI, C., GILBERT, G., CARN, H., GARNIER, S. & MOORE, B. C. J. 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proceedings of the National Academy of Sciences of the United States of America, 103**,** 18866-18869.

MEDDIS, R. & HEWITT, M. J. 1991a. Virtual Pitch and Phase Sensitivity of a Computer-Model of the Auditory Periphery .1. Pitch Identification. Journal of the Acoustical Society of America, 89**,** 2866-2882.

MEDDIS, R. & HEWITT, M. J. 1991b. Virtual Pitch and Phase Sensitivity of a Computer-Model of the Auditory Periphery .2. Phase Sensitivity. Journal of the Acoustical Society of America, 89**,** 2883-2894.

MEDDIS, R. & HEWITT, M. J. 1992. Modeling the identification of concurrent vowels with different fundamental frequencies. Journal of the Acoustical Society of America, 91**,** 233-245.

MEDDIS, R. & OMARD, L. 1997. A unitary model of pitch perception. Journal of the Acoustical Society of America, 102**,** 1811-1820.

MICHEYL, C., BERNSTEIN, J. G. W. & OXENHAM, A. J. 2006. Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise. Journal of the Acoustical Society of America, 120**,** 1493-1505.

MOORE, B. C. J., PETERS, R. W. & STONE, M. A. 1999. Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. Journal of the Acoustical Society of America, 105**,** 400-411.

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. Hearing Research, 24**,** 1-15.

PATTERSON, R. D. 1994. The sound of a sinusoid: Time-interval models. The Journal of the Acoustical Society of America, 96**,** 1419-1428.

PATTERSON, R. D., ALLERHAND, M. H. & GIGUERE, C. 1995. Time-Domain Modeling of Peripheral Auditory Processing - a Modular Architecture and a Software Platform. Journal of the Acoustical Society of America, 98**,** 1890-1894.

PATTERSON, R. D. & IRINO, T. 1998. Modeling temporal asymmetry in the auditory system. The Journal of the Acoustical Society of America, 104**,** 2967-2979.

PLACK, C. J. & CARLYON, R. P. 1995. Differences in Frequency-Modulation Detection and Fundamental-Frequency Discrimination between Complex Tones Consisting of Resolved and Unresolved Harmonics. Journal of the Acoustical Society of America, 98**,** 1355-1364.

PLACK, C. J. & WHITE, L. J. 2000a. Perceived continuity and pitch perception. The Journal of the Acoustical Society of America, 108**,** 1162-1169.

PLACK, C. J. & WHITE, L. J. 2000b. Pitch matches between unresolved complex tones differing by a single interpulse interval. The Journal of the Acoustical Society of America, 108**,** 696-705.

PLOMP, R. 1964. Rate of Decay of Auditory Sensation. The Journal of the Acoustical Society of America, 36**,** 277-282.

POLLACK, I. & TRITTIPOE, W. J. 1959. Binaural Listening and Interaural Noise Cross Correlation. The Journal of the Acoustical Society of America, 31**,** 1250-1252.

PRESSNITZER, D., PATTERSON, R. D. & KRUMBHOLZ, K. 2001. The lower limit of melodic pitch. The Journal of the Acoustical Society of America, 109**,** 2074-2084.

ROSS, B., TREMBLAY, K. L. & PICTON, T. W. 2007. Physiological detection of interaural phase differences. The Journal of the Acoustical Society of America, 121**,** 1017-1027.

SACHS, M. B. & KIANG, N. Y. S. 1968. Two-Tone Inhibition in Auditory-Nerve Fibers. The Journal of the Acoustical Society of America, 43**,** 1120-1128.

SCHEFFERS, M. T. M. 1983. Sifting Vowels: Auditory Pitch Analysis and Sound Segregation. PhD, Rijkssuniversitet te Groningen.

SCHIANO, J. & TRAHIOTIS, C. 1985. Lateralization of low-frequency tones and narrow bands of noise. The Journal of the Acoustical Society of America, 77**,** 1563-1570.

SCHOUTEN, J. F., RITSMA, R. J. & CARDOZO, B. L. 1962. Pitch of the Residue. The Journal of the Acoustical Society of America, 34**,** 1418-1424.

SHACKLETON, T. M. & CARLYON, R. P. 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. The Journal of the Acoustical Society of America, 95**,** 3529-3540.

SHEFT, S. & YOST, W. A. 1990. Temporal integration in amplitude modulation detection. The Journal of the Acoustical Society of America, 88**,** 796-805.

SUMMERFIELD, Q. & ASSMANN, P. F. 1991. Perception of Concurrent Vowels - Effects of Harmonic Misalignment and Pitch-Period Asynchrony. Journal of the Acoustical Society of America, 89**,** 1364-1377.

TERHARDT, E. 1974. Pitch, Consonance, and Harmony. Journal of the Acoustical Society of America, 55**,** 1061-1069.

VIEMEISTER, N. & PLACK, C. 1993. Time analysis. In: W.A. Yost, A.N. Popper, and R.R. Fay (Eds.) Human psychophysics, 3, (Springer-Verlag, New York) 116–154.

VIEMEISTER, N. F. 1979. Temporal modulation transfer functions based upon modulation thresholds. The Journal of the Acoustical Society of America, 66**,** 1364-1380.

VIEMEISTER, N. F. & WAKEFIELD, G. H. 1991. Temporal Integration And Multiple Looks. Journal of the Acoustical Society of America, 90**,** 858-865.

WALTERS, T. C. 2010. Auditory Processing of Communication Sounds. PhD, University of Cambridge.

WHITE, L. J. & PLACK, C. J. 1998. Temporal processing of the pitch of complex tones. Journal of the Acoustical Society of America, 103**,** 2051-2063.

WHITE, L. J. & PLACK, C. J. 2003. Factors affecting the duration effect in pitch perception for unresolved complex tones. Journal of the Acoustical Society of America, 114**,** 3309-3316.

WIEGREBE, L. 2001. Searching for the time constant of neural pitch extraction. Journal of the Acoustical Society of America, 109**,** 1082-1091.

WIEGREBE, L., PATTERSON, R. D., DEMANY, L. & CARLYON, R. P. 1998. Temporal dynamics of pitch strength in regular interval noises. Journal of the Acoustical Society of America, 104**,** 2307-2313.

WIGHTMAN, F. L. 1973. Pattern-Transformation Model Of Pitch. Journal of the Acoustical Society of America, 54**,** 407-416.

YOST, W. A. 1982. The dominance region and ripple noise pitch: A test of the peripheral weighting model. The Journal of the Acoustical Society of America, 72**,** 416-425.

YOST, W. A. 1996. Pitch strength of iterated rippled noise. Journal of the Acoustical Society of America, 100**,** 3329-3335.

YOST, W. A. & HILL, R. 1978. Strength of the pitches associated with ripple noise. The Journal of the Acoustical Society of America, 64**,** 485-492.

YOST, W. A., PATTERSON, R. & SHEFT, S. 1996. A time domain description for the pitch strength of iterated rippled noise. Journal of the Acoustical Society of America, 99**,** 1066-1078.

ZWICKER, U. T. 1984. Auditory recognition of diotic and dichotic vowel pairs. Speech Communication, 3**,** 265-277.

ZWISLOCKI, J. & FELDMAN, R. S. 1956. Just Noticeable Differences in Dichotic Phase. The Journal of the Acoustical Society of America, 28**,** 860-864.