

Published in final edited form as:

J Acoust Soc Am. 2008 February ; 123(2): 973–985. doi:10.1121/1.2821986.

Behavioral and physiological correlates of temporal pitch perception in electric and acoustic hearing

Robert P. Carlyon^{1,a}, Suresh Mahendran¹, John M. Deeks¹, Christopher J. Long¹, Patrick Axon³, David Baguley³, Stefan Bleeck^{2,b}, and Ian M. Winter²

¹ MRC Cognition & Brain Sciences Unit, 15 Chaucer Rd., Cambridge CB2 2EF, England ² Physiological Laboratory, University of Cambridge, Downing St., Cambridge CB2 3EG, England ³ Addenbrooke's NHS Trust, Hills Rd, Cambridge, England

Abstract

In the “4-6” condition of experiment 1, normal-hearing (NH) listeners compared the pitch of a bandpass-filtered pulse train, whose inter-pulse intervals (IPIs) alternated between 4 and 6 ms, to that of isochronous pulse trains. Consistent with previous results obtained at a lower signal level, the pitch of the 4-6 stimulus corresponded to that of an isochronous pulse train having a period of 5.7 ms – longer than the mean IPI of 5 ms. In other conditions the IPI alternated between 3.5-5.5 ms and 4.5-6.5 ms. Experiment 2 was similar but presented electric pulse trains to one channel of a CI. In both cases, as overall IPI increased, the pitch of the alternating-interval stimulus approached that of an isochronous train having a period equal to the mean IPI. Experiment 3 measured compound action potentials (CAPs) to alternating-interval stimuli in guinea pigs and in NH listeners. The CAPs to pulses occurring after 4-ms intervals were smaller than responses to pulses occurring after 6-ms intervals, resulting in a modulated pattern that was independent of overall level. The results are compared to the predictions of a simple model incorporating auditory-nerve (AN) refractoriness, and where pitch is estimated from 1st-order intervals in the AN response.

I. INTRODUCTION

A. Background

In normal, acoustic hearing, the pitch of a complex tone is dominated by the lower-numbered harmonics, which are resolved by the peripheral auditory system (Plomp, 1967; 1985). The reasons for this domination remain unclear, but may arise from one or more of the following: (i) the presence of place-of-excitation cues (ii) the existence of a “match” between the place-of-excitation and the frequency of phase locking (Moore, 1989; Oxenham et al., 2004; Moore and Carlyon, 2005), (iii) superior phase-locking to fine-structure than to the envelope (Moore et al., 2006), and (iv) differences in the relative timing of the responses of different auditory nerve (AN) fibers (Loeb et al., 1983; Shamma, 1985; Loeb, 2005; Moore and Carlyon, 2005). These timing differences are reflected by a steep, level-independent transition in the function relating the phase of AN firing to place along the cochlea, and the place at which it occurs may code the frequency of pure tones or of

Corresponding author: Robert Carlyon, Tel +44 1223 273717, Fax: +44 1223 359062.

^aElectronic mail: bob.carlyon@mrc-cbu.cam.ac.uk

^bCurrent address: Institute of Sound and Vibration Research, University of Southampton, England

PACS numbers: 43.66Hg, 43.66Ts, 43.64Pg

resolved harmonics (Kim et al., 1980; Loeb et al., 1983; Shamma, 1985; Loeb, 2005; Moore and Carlyon, 2005).

The present article investigates, instead, pitches that can only be conveyed by the temporal response of AN fibers tuned to the frequency components of the stimulus. Stimuli that elicit this “purely temporal” pitch can be produced by bandpass filtering an acoustic pulse train so that it contains only high-numbered, unresolved harmonics, or by presenting an electric pulse train to one channel of a cochlear implant (“CI”: McKay and Carlyon, 1999; Carlyon et al., 2002; van Wieringen et al., 2003). As we have pointed out before (Carlyon et al., 2002), such stimuli, although producing a fairly weak pitch, are of both theoretical and practical interest. First, by restricting the number of peripheral cues available, they can provide a simple test of more general models of pitch perception. Second, cochlear implants encode fundamental frequency (F0) using this purely temporal code, and understanding it may provide a basis for improving the generally poor pitch percepts experienced by CI users (McDermott, 1997; Moore and Carlyon, 2005). Third, the extent to which similar patterns of results can be obtained with acoustic and electric stimuli may allow one to develop an accurate simulation of CI hearing using NH listeners. Such simulations could, for example, be useful in the development of novel signal-processing strategies and/or experimental procedures.

Another feature of the stimuli used to study purely temporal pitch perception is that a comparison of behavioral data with the response of the AN is much more straightforward than when resolved harmonics are present. For acoustic pulse trains lacking resolved harmonics, one can ignore place-of excitation cues, and it is likely that there is no consistent cue conveyed by the relative timing of the responses of different AN fibers: for example, the steep phase transition observed across the AN fiber array for resolved partials (Kim et al., 1980) is absent, and all fibers that do respond to each pulse do so at approximately the same time (Carlyon and Shamma, 2003). Hence, a good estimate of the information conveyed by the AN can be obtained from the whole-nerve response to each pulse, as measured by the compound action potential (CAP). We exploit this fact to compare AN and behavioral responses to very similar stimuli. Such a technique would not be possible when stimuli consist of resolved harmonics; the temporal smoothing produced by peripheral filtering would prevent the measurement of CAPs throughout the stimulus, and potentially important information on place-of-excitation and on the relative timing of AN responses would be lost. To the extent that the behavioral results are similar for NH and CI listeners, this method also provides an indirect way of studying physiological correlates of temporal pitch perception in electric hearing.

The behavioral experiments that we report here exploit and extend a paradigm previously described by Carlyon et al (2002). They asked both NH listeners and CI users to compare the pitch of a pulse train whose IPIs alternated between 4 and 6 ms (Fig. 1a) with that of a range of isochronous pulse trains, in which the IPI was constant throughout the stimulus. They found very similar results with the two groups of listeners: the pitch of the “4-6” stimulus corresponded to that of an isochronous train having an IPI of about 5.7 ms. They noted that this match was longer than the mean interval (5 ms) of the “4-6” stimulus, and shorter than its 10-ms period. They proposed a model that could account for the pitch of these stimuli, and also for the results obtained by themselves and others using several different paradigms (Carlyon, 1997; Plack and White, 2000; Carlyon et al., 2002). The model assumed that only the first-order intervals in the stimulus determined pitch, and that longer intervals received greater “weights” than shorter ones.

Carlyon et al's (2002) model successfully accounted for a wide range of findings using a single set of parameters. Furthermore, the idea that the pitches of pulse trains are dominated

by first-order intervals is consistent with the conclusions from a number of other recent studies (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001; Yost et al., 2005). However, it has a number of limitations. One of these is illustrated by a study by McKay and Carlyon (1999). They presented both NH and CI listeners with a set of pulse trains, each of which was amplitude modulated by increasing the level of every n th pulse. Fig. 1b illustrates the fact that such stimuli can be characterised as having a carrier rate (R_c) and a modulator rate (R_m). By performing a multi-dimensional scaling (MDS) experiment using stimuli having different combinations of R_m and R_c , McKay and Carlyon showed that listeners were sensitive to both the carrier and modulation rates. Carlyon et al.'s (2002) model, however, produces a single pitch value, and cannot account for this finding. Indeed, the fact that subjects are at all sensitive to the modulator rate implies that pitch cannot be entirely determined by first-order intervals in the *stimulus*, at least when some pulses have a higher amplitude than others. Not surprisingly, when subjects are forced to match the pitch of a modulated pulse train, then, as the modulation depth increases from 0 to 100%, the match decreases from the carrier to the modulator rate (McKay et al., 1995; McKay and Carlyon, 1999).

Here, we investigate the hypothesis that a similar phenomenon can occur for equal-amplitude pulse trains, provided that the *auditory nerve* response, as measured by the CAP, is amplitude modulated. Specifically, for a “4-6” pulse train, the response to pulses occurring after a 4-ms interval may be smaller than that to those occurring after a 6-ms interval (Fig. 1c). A simple model, incorporating this effect, would predict the pitch as follows: If there exists an array of more-central neurons that fire only when the CAP amplitude exceeds a given threshold (e.g., when a threshold number of AN fibers fire synchronously), and if some of those neurons have thresholds higher than the response to pulses occurring after 4-ms intervals (Fig. 1c, dashed line), then those neurons will fire every 10 ms. The remainder, having thresholds lower than this criterion, will fire after every pulse. Pitch matches may then be obtained by a simple average of the first-order intervals in the responses of these two sets of more-central neurons. For the 4-6 stimulus, this would be a combination of 4 and 6 ms intervals (lower-threshold neurons) and of 10 ms intervals (higher-threshold neurons). This general scheme would be consistent with the results of van Wieringen et al (2003), who used alternating-interval pulse trains in which the pulses after either the short or the long intervals were attenuated (Figs. 1d, 1e). They found that pitch was lower when the pulses after the shorter intervals were attenuated (Fig. 1d), consistent with the neural response to each lower-amplitude pulse occurring after short IPIs being further attenuated by refractory effects originating from the previous, higher-amplitude pulse. This pitch difference decreased with increasing overall IPI (e.g. from 4-6 ms to 8-12 ms), both in NH and CI listeners, consistent with an explanation based on refractoriness, if one further assumes that the recovery function starts to flatten off over this range of delays. This general scheme would also be consistent with McKay and Carlyon's (1999) finding that listeners can perceive both the carrier and modulator rates of AM stimuli, if one assumes that they can selectively “attend” to different subsets of the more-central neurons. At the same time, it would also be consistent with reports that higher-order intervals do not have a large effect on the pitches of pulse trains that do not produce large and/or regular modulations in the AN response (Kaernbach and Demany, 1998; Plack and White, 2000; Kaernbach and Bering, 2001; Yost et al., 2005).

Two further points are worth making. First, Pressnitzer et al (2001; 2004) have argued that higher-order intervals between *pulses* can be transformed into 1st-order intervals between *spikes* in the response of AN and cochlear-nucleus neurons. As with the simple model proposed here, their research emphasizes the need to consider the input to the pitch mechanism in terms of neural activity, rather than solely considering the statistics of the stimulus. Second, we should stress that our simple model assumes that a summary statistic is

derived from the summed response of multiple fibers, rather than assuming that a statistic (such as the inter-spike-interval histogram; “ISIH”) is derived from each fiber, with these individual statistics then being combined. This distinction is important, because it has been argued, based on the results of single-unit recordings (Cariani and Delgutte, 1996), that a temporal code based on first-order intervals should depend strongly on overall level. However, Carlyon et al (2002) have argued that this should not necessarily be the case when the responses of several neurons are combined before a summary statistic is derived.

B. Overview of model and experiments

The aim of the experiments described here was to compare behavioral measures of temporal pitch perception in NH and CI users to the neural response as measured by the CAP. In particular, we wished to determine the extent to which the behavioral measures could be accounted for by the type of simple neural model described in the previous subsection. Specifically, we compare the results to the predictions of a model in which the thresholds of the “more central” neurons are uniformly distributed across level, and where the predicted pitch is obtained from an unweighted sum of the 1st-order intervals at the output of this more-central population. For example, if the CAP to a 4-6 pulse train were amplitude modulated by 10%, then 10% of the “more central” neurons would fire every 10 ms, and 90% would follow the “4-6” pattern. The period corresponding to the pitch would then be $(0.1 \times 10) + (0.45 \times 4) + (0.45 \times 6) = 5.5$ ms.

The model described above assumes that the CAP provides an accurate measure of the whole-nerve input to the more-central neural population, and that the time window over which this input is summed corresponds simply to the smoothing inherent in the CAP measurement, which likely derives from the integrative properties of the inner-hair-cell and AN-fiber membranes. As noted in the Introduction, this approach allows a direct comparison of behavioral results to physiological measures obtained using very similar stimuli under conditions where, unlike the case where resolved harmonics are present, the neural code is limited to purely temporal cues.

A “first pass” test of the neural model is, of course, that the CAP response to an alternating-interval pulse train is indeed amplitude modulated. To test this, experiment 3a measured CAPs to bandpass filtered “4-6” pulse trains in five anesthetized guinea pigs (GPs) and from two NH listeners. The results showed that the predicted form of modulation was indeed present. A more stringent test comes from the requirement that pitch be largely level-independent. As noted above, we have previously argued that a statistic that is derived from the responses of multiple fibers may be robust to changes in overall level, and this feature would be needed to avoid the prediction that pitch changed markedly with level. The results of experiment 3a showed that the modulation in the CAP to 4-6 pulse trains was indeed largely independent of level over a 50-dB range. In addition, the results of experiment 1 showed that the pitch of 4-6 stimuli was judged by NH listeners to be similar to that of an isochronous pulse train having a period of about 5.6 ms, a result very similar to that obtained previously at a 24-dB lower level (Carlyon et al., 2002).

A yet more demanding test of the model is that, as stimulus parameters are manipulated, the behavioral results should quantitatively follow the predictions. Experiment 1 also included two new conditions in which the 4-6 pulse train was replaced by one with slightly shorter (“3.5-5.5”) or longer (“4.5-6.5”) intervals. These conditions were originally included because we wished to measure CAPs to alternating-interval pulse trains, and wanted to avoid stimuli whose F0 (which was 100 Hz for the 4-6 train) was a harmonic of the 50-Hz U.K. mains frequency. The inclusion of such stimuli also led to an interesting prediction. One would, of course, expect the matched pitch to correspond to longer intervals for a 4.5-6.5 train than for a 3.5-5.5 train. In addition, if the *relative* change in refractoriness

between the long and the short intervals is smaller at longer overall IPIs, then one might also expect a decrease in the proportion of central neurons responding only to every other pulse. This in turn would cause the matched pitch to be closer to the mean of the two intervals in the alternating-interval stimulus; that is, closer to 5.5 ms for the 4.5-6.5 stimulus than to 4.5 ms for the 3.5-5.5 stimulus. This hypothesis was also tested in experiment 2 with CI listeners, using 4-6 and 5-7 pulse trains. The hypothesis was confirmed behaviorally with both groups of listener. However, the pattern of results was not reflected by differences in the modulation depth of the CAP response to the 3.5-5.5, 4-6, and 4.5-6.5 stimuli, obtained in Experiment 3b from three additional GPs. We conclude that, although the general form of model described here can account qualitatively for a wide range of data, the physiological data obtained from the GP AN does not account for the effects of varying inter-pulse interval. Two possible explanations for this discrepancy – species differences and the existence of an additional source of refractoriness – are discussed, and, in the latter case, a quantitative estimate of the additional refractoriness needed is presented. The aim of all these studies was not to disprove Carlyon et al's (2002) model, but rather to see whether the refractory properties of the auditory nerve would allow a more simple explanation of the data that would dispense with the need for a central weighting function.

II. EXPERIMENT 1: TEMPORAL PITCH STUDIED WITH NH LISTENERS

A. Method

All pulse trains were generated digitally in the time domain and played out through a 16-bit DAC (CED1401 *plus* laboratory interface) at a sampling rate of 50,000 Hz. They were then passed through an antialiasing filter (Kemo VBF25.01; 100 dB/octave) and bandpass filtered between 3900-5400 Hz using a lowpass and a highpass 8th-order Butterworth filter in series (Kemo VBF25.03; 48 dB/octave). The duration of each pulse train was 400 ms, including 10-ms raised-cosine ramps. The level of every pulse train was 78 dB SPL. This was higher than the 54 dB SPL used by Carlyon et al (2002), in order to aid comparison with the CAPs to the same stimuli in experiment 3, for which a higher level was considered desirable in order to obtain a more robust neural response. Pulse trains were then attenuated (Tucker-Davis Technologies PA2) and mixed with a pink noise. The noise was gated on and off synchronously with the pulse train and was played out of a second DAC. A fresh 400-ms sample of noise was selected for each presentation by sampling from a random point in a previously generated 2-sec wave file (CoolEdit 2000, Syntrillium software Inc). The noise was bandpass filtered between 100-3900 Hz (Kemo VBF25.03 highpass and lowpass filters in series; attenuation 48 dB/octave), attenuated (TDT PA2), and mixed with the pulse train. Its spectrum level at 1000 Hz was 42 dB SPL. Stimuli were then presented via one earpiece of a Sennheiser HD250 headset to a listener seated in a double-walled sound-attenuating booth. Calibration was performed with the aid of a B&K type 4153 artificial ear and an HP3561A spectrum analyser.

There were three conditions, defined by the durations of the IPIs in the alternating-interval stimulus: 3.5-5.5 ms, 4-6 ms, and 4.5-6.5 ms. In each trial of each condition, the listener heard the alternating-interval stimulus and one of five isochronous pulse trains, presented in random order. The IPIs for the isochronous trains in the 3.5-5.5 condition were 2.5, 3.5, 4.5, 5.5, and 6.5 ms. In the 4-6 and 4.5-6.5 conditions these values were increased by 0.5 and 1 ms, respectively. The listener was instructed to identify which of the two stimuli in the trial had the higher pitch by clicking on one of two virtual buttons on a computer monitor. No feedback was provided. Each listener performed 10 repeats ($\times 5$ isochronous stimuli = 50 trials) for each condition before moving on to the next one. The standard stimulus started with the shorter of its two IPIs for five of these repeats and with the longer IPI for the other five. No differences were observed between the results obtained with these two types of trial, and they were therefore averaged. Each condition was then repeated in the same order

until each listener had completed 200 trials per data point, with the exception of listeners NH2 and NH7, who completed 150 and 160 trials respectively. Seven NH listeners participated in the experiment.

B. Results

Each psychometric function in Fig. 2 shows the proportion of trials, averaged across listeners, on which each isochronous pulse train was judged higher in pitch than the alternating-interval stimulus in one condition. It can be seen that, for each alternating-interval stimulus, the psychometric function spans the range from below 20% to above 95%. The functions for stimuli with longer IPIs (e.g. 4.5-6.5) are to the right of those with shorter IPIs (e.g. 3.5-5.5), showing that their perceived pitch corresponded to a longer IPI, and was therefore lower. To estimate the period of an isochronous stimulus judged equal in pitch to each standard, the psychometric function for each subject and condition was subjected to a probit analysis. The point of subjective equality ("PSE"), corresponding to the point at which the fitted function passed through 50% intercept on the ordinate, is shown for each subject and condition in Fig. 3a; mean data are shown by the thick dashed curve with square symbols. Not surprisingly, the PSE increases as the IPIs in the alternating-interval standards increase from 3.5-4.5 through 4-6 to 4.5-6.5 ms. For the 4-6 stimulus, the mean PSE is 5.64 ms, close to the 5.7-ms reported by Carlyon et al (2002) for stimuli presented at a softer overall level (54 vs. 78 dB SPL). Fig. 3b shows the PSE in each condition divided by the average IPI in the standard for that condition. In the absence of refractory effects, our simple, unweighted model, would predict a ratio of one. The ratio decreases from 1.18 for the 3.5-5.5 stimulus to 1.10 for the 4.5-6.5 stimulus, consistent with the change in refractoriness between 4.5 and 6.5 ms being smaller than that between 3.5 and 5.5 ms. This trend was confirmed by a one-way ANOVA ($F(2,10)=15.8$, $p<0.001$, Huynh-Feldt sphericity correction). We should also note that, when averaged across listeners, the results are consistent with Carlyon et al's (2002) model, the predictions of which are shown by the heavy dashed line with "plus" symbols in Figs. 3a and 3b. These lines are superimposed on the mean data (dashed line, squares), reflecting the good fit of the model. The solid circles without lines will be discussed in section V.A.

III. EXPERIMENT 2: TEMPORAL PITCH STUDIED WITH CI USERS

A, Method

The method used for experiment 2 was generally similar to that for experiment 1. An important difference is that, instead of presenting filtered acoustic pulse trains to NH listeners, we presented electric pulse trains via a bipolar pair of intracochlear electrodes of a CI. Five listeners took part, all of whom had been implanted with either the CI22 or CI24 device manufactured by Cochlear Corp. The listeners' details, including information on the device used by each of them, are given in Table 1. All stimulation was on electrode 17, with electrode 13 serving as the return electrode. This corresponds to so-called "BP+3" mode, with approximately 3 mm between electrodes. Stimuli consisted of 400-ms trains of biphasic pulses, with each pulse having a phase duration of 100 μ s and an inter-phase gap of 8 μ s. Standard stimuli consisted of "4-6" and "5-7" pulse trains. The isochronous stimuli to be compared to the 4-6 standard had IPIs of 3,4,5,6, and 7 ms; those to be compared to the 5-7 ms standard had IPIs of 4,5,6,7, and 8 ms.

At the start of the experiment, the threshold and most-comfortable ("C") level was obtained for the 4-6 stimulus for each subject, using the same electrodes and configuration as in the main experiment. The 5-7 standard was then loudness-balanced to the 4-6 standard using the procedure similar to that described by McKay & Carlyon (1999). One of the two stimuli was presented first, followed 500-ms later by the second stimulus at a level that was 10 clinical

current units (approx 1.76 dB) lower. The subject could then adjust the level of the second sound to be presented on the next trial, by pressing one of six virtual buttons (labeled ‘++’, ‘+’, ‘-’, ‘--’, ‘---’, and ‘---’) on a computer screen. This procedure continued until the subject was satisfied that the two stimuli had equal loudness. It was then repeated. The roles of the fixed and standard stimuli were then swapped, and the procedure repeated twice. The average difference between the levels of the two stimuli over these four runs was used to equate their loudness. Each standard was then loudness-balanced to the isochronous pulse trains having the longest and shortest IPIs to be compared to it (e.g. 3 and 7 ms for the 4-6 standard). Levels for isochronous stimuli having intermediate IPIs were obtained via linear interpolation in clinical current units (CUs), where one CU is equal to approximately 0.176 dB. This was deemed reasonable because loudness does not vary markedly with level over the range of IPIs studied here (McKay and McDermott, 1998; Carlyon et al., 2002).

B. Results

The PSEs for each condition were obtained in the same way as for experiment 1 and are shown in Fig. 4a. Consistent with the results of that experiment, we obtained the unsurprising finding that the PSE was longer on average for the 5-7 than for the 4-6 standard. More interesting is the fact that, as shown in Fig. 4b, the ratio of the PSE to the mean interval in the standard was lower for the 5-7 than for the 4-6 stimulus. This finding was obtained for four out of the five listeners, and was significant overall, as revealed by a paired-samples t-test ($df=4$, $p<0.05$, two-tailed). This result is consistent with that obtained in experiment 1, and with the idea that refractory effects influence temporal pitch perception in electric and acoustic hearing in a roughly similar way. Again, however, we should note that the results are also roughly consistent with the central weighting function proposed by Carlyon et al (2002) (heavy dashed line and ‘plus’ symbols). Presumably, however, any central mechanism will operate on the AN response rather than on the physical stimulus. The aim of the next experiment was to determine whether refractory properties of the AN would produce a peripheral representation that would allow one to dispense with the need for such a weighting function.

IV. EXPERIMENT 3: CAP MEASUREMENTS.

A. Subjects

Experiment 3a measured CAPs to the same stimuli in anesthetized guinea pigs (“GPs”) and in (human) NH listeners. The former group consisted of five pigmented GPs with weights between 330 and 585g and CAP thresholds within 5 dB of the norms obtained in author IMW’s laboratory, where the GP recordings were obtained. The latter group consisted of five normal-hearing adults, including subjects NH1, NH2, and NH3 from experiment 1. However, for reasons that will be discussed in section IV.D, it was only possible to record reliable responses to each pulse in a train from listeners NH1 and NH7. Experiment 3b measured CAPs to equal-amplitude and amplitude-modulated pulse trains from an additional three GPs.

B. Recording

1. GPs—The method of recording was as described by Neuert *et al.* Briefly, GPs were anesthetized with urethane [1.5 g/kg, intraperitoneally (ip)]. Hypnorm was administered as supplementary analgesia (1mg/kg, intramuscularly). i.m.). Anesthesia and analgesia were maintained at a depth sufficient to abolish the pedal-withdrawal reflex (front paw). Additional doses of Hypnorm (1 ml/kg) or urethane (1 ml) were administered on indication. Incisions were preinfiltrated subcutaneously with the local anesthetic Lignocaine (Norbrook Laboratories, Newry, UK). Core temperature was monitored with a rectal probe and maintained at 37°C using a thermostatically controlled heating blanket (Harvard Apparatus,

Holliston, MA). The trachea was cannulated, and the animal was ventilated artificially with a pump if it showed signs of suppressed respiration. Surgical preparation and recordings took place in a sound-attenuated chamber (Industrial Acoustics). The animal was placed in a stereotaxic frame that had ear bars coupled to hollow speculae designed for the guinea pig ear. A midsagittal scalp incision was made, and the periosteum and the muscles attached to the temporal and occipital bones were removed. The bone overlaying the left bulla was fenestrated, and a silver-coated wire was placed on the round window of the cochlea to record the CAPs. The hole was resealed with petroleum jelly. The wire electrode was connected via an amplifier (WPI DAM 50, gain = $\times 10\,000$) to an interface box (Hammerfall DSP multiface MIDI 24 Bit 96kHz Multichannel interface), and then stored on a PC via an interface card (RME Intelligent Audio Solutions Hammerfall DSP system-HDSP cardbus interface), for off-line analysis. All GP experiments were performed in accordance with terms and conditions of the project licence issued by the U.K. Home Office to author IMW.

2. Human NH listeners—Pre-test examination and placement of electrodes for the human subjects was similar to that used in standard clinical electrocochleography. The active (recording) electrode was a soft-tipped electrode (Bio-logic Systems Corp TM-ECochGtrode) placed gently on the surface of the tympanic membrane. Along with single-use common (forehead) and reference (contralateral mastoid) electrodes (SLE diagnostics, ref. M0872) it was connected to one channel of the head-box of a Digitimer D360 8-channel patient amplifier (Digitimer Ltd., U.K.). The Digitimer amplifier was set to a passband of 70-1500 Hz, a gain of 50 000, and the notch filter turned on to eliminate 50 Hz hum (U.K. mains frequency). The output of the amplifier was then connected to the interface box, and the results stored on the PC for off-line analysis. For safety reasons, all equipment with the exception of the patient amplifier was battery-powered.

C. Stimulus Generation

The method of stimulus generation was similar for both groups of subjects. In experiment 3a it was identical to that described in experiment 1 with the following exceptions: (i) Filtering was carried out in software, using 8th-order lowpass and highpass Butterworth filters, in series (ii) The sampling rate was 96 kHz (iii) the duration of each pulse train was 100 ms (iv) There were no onset and offset ramps, and (v) No pink noise was presented. The pink noise was used in experiment 1 to prevent listeners from using cochlear distortion products, which we did not expect to affect the CAP. Only the 4-6 and 6-4 pulse trains were presented. The stimulus level was 78 dB SPL for the NH subjects. For the GPs, a range of levels between 38 and 88 dB SPL, in 10-dB steps, were tested for the 4-6 stimulus. Additionally, measurements for the 6-4 stimulus were obtained at 88 dB SPL for all five GPs, and at 78 dB SPL for GP2, GP3, and GP4.

In experiment 3b, CAPs were obtained in three further GPs for an additional set of stimuli, presented at a level of 78 dB SPL. These consisted of equal amplitude pulse trains in which the odd- and even-numbered intervals, in ms, were 3.5-5.5, 5.5-3.5, 4-6, 6-4, 4.5-6.5, 6.5-4.5, 8-12, and 12-8. In addition, CAPs were obtained for 4-6 and 8-12 pulse trains in which the pulses occurring after the short (4 or 8 ms) or long (6, 12 ms) intervals were attenuated by 2 or 6 dB (Fig. 1d,e). These amplitude-modulated pulse trains, which resemble a subset of those used by van Wieringen et al (2003) are designated here by the letter “S” (for “short”) or “L” (for “long”), indicating the intervals after which pulses were attenuated, followed, optionally, by a number indicating the attenuation in dB. (For example, 4-6S2 is a 4-6 stimulus in which all pulses after the shorter (4-ms) interval are attenuated by 2 dB). The final number is omitted whenever we refer to a general class of stimulus without specifying the amount of attenuation used.

Waveform files were transferred from a PC to the same interface box as used for response collection. For the GPs they were played out via a power amplifier (Rotel RB971 Mk 2) and a custom-built end attenuator before being presented over a speaker (30-1777 tweeter: RadioShack, FortWorth, Texas) that was mounted in a coupler designed for the GP ear. Stimuli were acoustically monitored using a condenser microphone (B&K 4134) attached to a calibrated 1-mm-diameter probe tube that was inserted into the speculum, close to the eardrum. For NH subjects they were presented via a headphone amplifier to one earpiece of an Etymotic ER-3 insert phone. Because the extra-tympanic electrode used for the normal-hearing human subjects was likely to reduce the sound pressure level at the eardrum, the following procedure was adopted to correct for this. Detection thresholds for a 5-kHz pure tone in quiet were measured using a two-interval forced-choice task and the adaptive procedure described by Levitt (1971). Two adaptive runs were obtained and the results averaged. This procedure was performed before and after placement of the electrode, using Sennheiser HD-250 headphones, and the difference between the thresholds obtained (mean=9.4, s.e.=1.7 dB) was used as an estimate of the attenuation produced by the electrode. The sound pressure level delivered by the insert earphone during the CAP recordings was then increased to compensate for this attenuation.

In each recording run, 100-ms presentations of a given pulse train were presented repeatedly, with a 50-ms silent interval between bursts. Every other burst was inverted in polarity, in order to reduce the influence of stimulus artefact and of cochlear microphonics when the responses were averaged. Typically we averaged responses to 600 stimuli for each GP and to 2000 or 3000 stimuli for the NH listeners. For the NH listeners, several additional conditions were also run. Subject NH1's responses were measured under four conditions: the 100-ms 4-6 and 6-4 pulse trains used for the GPs, and, for reasons unrelated to the present study, the same two pulse trains with a 90-ms duration. 2000 responses were obtained in each condition. Subject NH7 was tested with the 100-ms 4-6 and 6-4 stimuli, and with the interval between pulse trains increased to 300 ms. 3000 responses were obtained in each condition.

D. Results

1. Experiment 3a: GPs—CAPs—CAPs obtained from GPs were very similar across animals. Fig. 5a shows the CAP to the first pulse in a 4-6 train, obtained from one GP. It shows the typical form (e.g. Murnane et al., 1998) consisting of a negative deflection followed by a positive deflection. Here, the amplitude of the CAP is defined as the difference, in μV , between these negative and positive peaks.

Fig. 5b shows the response from the same GP to an entire 100-ms 4-6 pulse train at a level of 78 dB SPL. It shows an alternating pattern of 4- and 6-ms intervals, reflecting the IPIs present in the stimulus. The overall amplitude of the response decreases rapidly over the first three whole periods (30 ms), before reaching an asymptote (c.f. Eggermont and Spoor, 1973). Averaged across GPs, the CAP amplitudes after 10, 20, and 30 ms were 65, 58, and 56% of that to the first pulse. The response to the last pulse in the stimulus had an amplitude that was 55% of that to the first.

Fig 5c shows a close-up of the response shown in Fig. 5b, focusing on the positive deflections between 30 and 80 ms. It can be seen that the responses after 4-ms intervals are smaller than those after 6-ms intervals. To quantify this difference whilst minimising the influence of short-term adaptation, we first excluded the responses to the first 30 ms of the pulse train¹. We then separately averaged the response amplitudes of all pulses occurring after 4-ms and after 6-ms intervals. At 78 dB SPL the average response after 6-ms intervals was 11.0% (s.e.=1.4%) greater than that after 4-ms intervals, a difference that was statistically significant (t-test, df=4, $p<0.05$). Fig. 6a shows that this ratio varied only

between 7% and 11% over the range 38-88 dB SPL. The level-independence of the AN response is illustrated further in Figs b) and c), which show the CAP waveforms obtained from GP2 at levels of 88 and 38 dB SPL respectively. Although the CAP is smaller and slightly noisier at the lower stimulus amplitude, both the general form of the waveform and the amount of amplitude modulation are similar at the two levels.

2. Experiment 3a: NH Human listeners—Fig. 7a shows the response to a single pulse in subject NH2. Apart from the smaller amplitude, the CAP is similar to that obtained from GPs. A similar response was also obtained in the other four human subjects. However, when we measured CAPs to the pulse trains, it was only for subjects NH1 and NH7 that we saw a clear response to each pulse. The reasons for this are discussed in a separate report (Mahendran et al., in press). One reason is that the response even to a single pulse is smaller than in the GPs, and adaptation both from previous pulse trains and throughout each pulse train may reduce the amplitude into the noise floor. Another is that, in the majority of subjects, the response to a single pulse was followed by a myogenic response having a latency sufficiently long (e.g. 20 ms) to be missed by the short time window over which CAPs are usually analysed in clinical practice. Mahendran *et al* suggested that the CAP to each pulse may be distorted by this longer-latency response to previous pulses. Neither subjects NH1 nor NH7 showed this long-latency response, which was also absent from our GP recordings.

NH7's CAPs to a 100-ms pulse train are shown in Fig. 7b. The response shows a series of CAPs following each pulse in the stimulus, as is further illustrated by the zoomed-in plot with gridlines in Fig. 7c. (Note that the vertical gridlines are separated by 4 and 6 ms, following the alternating intervals in the stimulus, and that the peaks in the CAP response are aligned with the gridlines). It can be seen that, as with the GPs, the response to pulses after 6-ms IPIs are larger than those to pulses after 4-ms IPIs. To quantify this difference, we averaged the responses after 4-ms and after 6-ms intervals in the same way as for the GPs. Averaged across the 4-6 and 6-4 stimuli, the responses after 6-ms intervals were 19.7% higher than those after 4-ms intervals. Rather than report a measure of inter-subject variability from only two subjects, we obtained a measure of the variability of the response within each subject. For subject NH7 this was done by, for the 4-6 and 6-4 trains separately, analysing the responses to pulses n , $n+5$, $n+10$ etc, and obtaining five separate measures corresponding to $n=1,2,3,4$ and 5. 3000 measures were obtained for each condition, so each sub-measure corresponded to the average of 600 responses. Analysing the data in this way, we obtained 95% confidence limits between 9.2 and 35.5%. Averaged across the four stimuli tested for subject NH1, responses after 6-ms intervals were 10.7% bigger than after 4-ms intervals. When the data from the four different stimuli were further sub-divided into 3 interleaved sets of 666 responses, we obtained 95% confidence limits between 2.1 and 16.6%. It is worth noting that the confidence limits for both subjects are wide but do not encompass any negative values.

It is clear that, even in the two NH subjects from whom we obtained meaningful data, the responses to pulses throughout a train are much noisier than the results obtained from GPs. The wide confidence intervals mean that, for the NH listeners, we cannot obtain an accurate measure of the degree of amplitude modulation in the AN response. What the results do

¹We waited until the effects of short-term adaptation had started to level off so that, when comparing the response to pulses occurring after 4-ms and 6-ms intervals, the results would not be overly influenced by the response to the first pulse that we analysed. To check that we had succeeded in doing so, we compared the ratio of the response amplitude after 6-ms vs. 4-ms intervals for the 4-6 pulse train and for the 6-4 pulse train in the three GPs for whom these data were available. The first pulse analysed would correspond to a 6-ms interval in the former case and to a 4-ms interval in the latter. If the measured ratio were overly influenced by the first pulse then it would therefore differ between these two stimuli. The two ratios were very similar (1.061 and 1.065, respectively), and did not differ significantly ($t(df=2)$, $p=0.1$).

demonstrate is that the same general finding applies *qualitatively* to GPs and NH human subjects.

3. Experiment 3b: GPs—Experiment 1 showed that, for human NH listeners, the pitch match for a 4.5-6.5 stimulus was closer to the average IPI in that stimulus than was the case for the 4-6 and 3.5-5.5 stimulus. In terms of our simple model, this would be consistent with there being less modulation in the AN response at longer overall IPIs. To test this, we measured the average CAP after long vs. short intervals for 3.5-5.5, 4-6, and 4.5-6.5 stimuli in three GPs. We also obtained measures for an “8-12” stimulus, in order to maximize the chances of seeing an effect of overall IPI. To minimize effects of short-term adaptation, data from the first three whole periods were excluded for all stimuli except 8-12, for which data from the first two periods (40 ms) were excluded. As in experiment 3a, these measures were also obtained for stimuli starting with the longer IPI (5.5-3.5, 6-4, 6.5-4.5, 12-8), and the results averaged. The ratio of the CAPs after the long vs. short intervals is shown for the three GPs in Table 2a. It can be seen that there is no tendency for the amount of modulation to decrease with increasing IPI over the range studied. Section V.A.3 discusses possible reasons for this discrepancy.

Table 2b shows the percentage difference in CAP amplitude for pulses after the 6-ms vs 4-ms intervals for a subset of the stimuli similar to those used by van Wieringen et al (2003). When the amount of attenuation is increased, the percentage CAP difference increases for the 4-6S stimuli, and decreases for the 4-6L stimuli. In the latter case, the difference becomes negative, reflecting the fact that the attenuation of pulses after the longer intervals overcomes the smaller refractory effects relative to pulses after the shorter intervals. Similar trends are seen for the 8-12 stimulus in Table 2c.

Although the data for the three GPs are quantitatively similar for the unmodulated stimuli (4-6 and 8-12), the effect of attenuating pulses after the longer or shorter intervals was much greater for GP6 than for the other two animals. We therefore restrict our discussion to describing two trends that were apparent in the data of all three GPs. First, as the modulation in the stimulus is increased from 2 to 6 dB, then, for both the 4-6S and 4-6L stimuli, the modulation in the CAP response also increases. This is consistent with van Wieringen *et al*'s finding that subjects matched to longer periods for stimuli with larger modulation depths. Second, the differential effect of attenuating pulses after the longer vs. the shorter intervals was greater for the 4-6 than for the 8-12 stimuli; a 3-way ANOVA (factors= 4-6 vs 8-12, attenuation amount, attenuation on long vs short) revealed a borderline interaction between the overall interval duration (4-6 vs 8-12) and whether the attenuation was applied to pulses after the longer vs. the shorter interval ($F(1,2)=16.03$, $p=0.057$).

V. DISCUSSION

A. Models of temporal pitch perception

1. Carlyon et al (2002)—Carlyon et al's 2002 model assumed that pitch was estimated using a weighted sum of 1st-order intervals in the stimulus. The model accounted for the fact that a 4-6 stimulus was matched in pitch to an isochronous sound having a period of 5.7 ms – longer than the 5-ms mean IPI in the 4-6 sound - by assuming that the weights increased with increasing IPI up to 10-12 ms. As shown in Fig. 3b, it can also account for the fact that this tendency to produce a match longer than the mean IPI is greater for the 3.5-5.5 than for the 4-6 stimulus, and smallest of all for the 4.5-6.5 pulse train. It does so because the difference in weights between IPIs of 3.5 and 5.5 ms is greater than that between 4 and 6 ms, which in turn is greater than that between 4.5 and 6.5 ms. The model could also account for the data of Plack and White (2000), who presented listeners with sequences of eight filtered pulses, with an IPI of 4 ms. They found that delaying the last four pulses, thereby increasing

one IPI in the stimulus, had a much bigger effect on pitch than was produced by advancing those pulses. The model succeeded because the increased IPI in the “delayed” stimulus received a smaller weight than the shortened IPI in the “advanced” stimulus.

Overall, Carlyon et al's 2002 model does a good job of predicting the pitch of equal-amplitude pulse trains, such as those used here and in previous experiments (Carlyon, 1997; Plack and White, 2000; Carlyon, 2002). However, as noted in the Introduction, it can account neither for the multiple pitches that can be heard in amplitude-modulated isochronous pulse trains (McKay and Carlyon, 1999), nor for the different effects of attenuating pulses occurring after the longer *vs* the shorter intervals in alternating-interval pulse trains (van Wieringen et al., 2003). Here we consider whether a simple model based on refractory properties of the auditory nerve can account for such data.

2. Neural model: general form and level independence—In the Introduction we described a new type of model in which an array of neurons, central to the AN, only respond when the amplitude of the CAP exceeds a certain fixed threshold value. Here we consider the simplest form of this scheme, in which the thresholds are distributed uniformly across the more-central neurons, and where pitch is estimated from an unweighted sum of the 1st-order intervals in the outputs of these neurons. One implementation of this idea could occur via an array of “synchrony detectors”, each responding when a threshold number of input fibers fire in synchrony. Our only assumptions concerning the time window over which synchrony detection occurs is that it includes those smoothing properties – e.g. integration by IHC and AN cell membranes – that are involved in the generation of the CAP, and that it is shorter than the shortest interval between any two successive CAPs described here (e.g., 3.5 ms).

Two aspects of the physiological data obtained in experiment 3 lend general support to the model. First, the CAPs to equal-amplitude, alternating-interval pulse trains are indeed amplitude modulated. Second, the depth of this modulation is largely independent across level. This second finding is important because the model assumes a uniform distribution of thresholds for the “more central” neurons, so that any marked change in modulation depth across level would predict a substantial change in pitch. In fact, as Fig. 6 shows, the modulation in the CAP response differs only from 7–11% over a 50 dB range of input levels. In terms of the model, this would produce matches that ranged only from 5.35 to 5.66 ms. This finding is also of more general theoretical importance as it helps resolve a potential paradox in the literature. A number of studies point to the conclusion that pitch is dominated by the 1st-order intervals in the stimulus, and that higher-order intervals have a smaller effect (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001; Carlyon, 2002; Yost et al., 2005). However, it has been argued that models of pitch that rely on 1st-order intervals is that such representations, when applied to the responses of single AN fibers, are highly level-dependent (Cariani and Delgutte, 1996; McKinney and Delgutte, 1999). The CAP measures obtained in experiment 3a support Carlyon et al's (2002) suggestion that this problem can be overcome if one assumes that the representation of 1st-order intervals is derived *after* the responses of individual AN fibers are combined. Our results and analysis also suggest that the “1st-order interval” approach should be modified such that, when the CAPs to some pulses are larger than those to others, intervals between these larger CAPs may contribute to pitch.

3. Neural model: Effect of inter-pulse interval—Our results also suggest, however, that the refractory properties of the AN, as processed by our simple model, cannot, by themselves, account for all aspects of temporal pitch perception. An important discrepancy can be seen in Fig. 3b; the downward slope of the line connecting the mean data reflects the fact that the pitch match to a 4.5–6.5 pulse train is closer to the mean IPI (5.5 ms) than is the

case for a 3.5-5.5 train (mean IPI= 4.5 ms). This would be consistent with the refractoriness model if the slope of the recovery function decreased over this range, so that the amount of modulation in the CAP waveform were smaller for the 4.5-6.5 than for the 3.5-5.5 stimulus. However, the GP recordings from experiment 3a did not reveal such a trend. To quantify this discrepancy we obtained an estimate of the function relating neural response to inter-pulse interval that would be necessary to account for the mean data shown in Figs 3a and 3b. The procedure adopted was as follows: (i) calculate the *relative* size of the CAPs to pulses occurring after the shorter *vs.* longer intervals that would be needed to account for the pitch data obtained with each of the 3.5-5.5, 4-6, and 4.5-6.5 stimuli. This gives three pairs of values, where the relative CAP amplitude for the two members of each pair are defined relative to each other. It does not constrain the relative amplitude between members of other pairs (e.g. between gaps of 3.5 and 4 ms) (ii) Assume that the CAP amplitude after a 3.5-ms interval has a value of 1, and that the amplitude after gaps of 4 and 4.5 ms are equal to $1+Add4$ and $1+Add4.5$ respectively. (iii) Assume that the form of the function relating CAP amplitude to inter-pulse interval (Δt) is $y=a.\ln(\Delta t-r)+b$. (iv) Adjust $Add4$, $Add4.5$, a , b , and r to minimize the least-squares error between this function and the data, using the routine “solver.exe” in Microsoft Excel, and with the constraint that $r \geq 0$.² The resulting function, $y=0.163\ln(\Delta t-2.5)+1$ is shown by the solid lines connecting squares in Fig. 8, with CAP amplitude plotted relative to that after a gap of 3.5 ms. It is initially steeper than a similar function fit to the *actual* data, obtained from the CAPs to the alternating-interval pulse trains in GPs ($y=0.267\ln(\Delta t)+0.656$; solid lines and triangles; see also Table 2), but, unlike the GP function, decreases in slope over the range studied.

One possible reason for the discrepancy stems from species differences. The dashed line in Fig. 8 shows the recovery functions obtained from the cat by Fitzpatrick et al (1999), defined as the probability of a single AN fiber firing in response to the second of two acoustic clicks, as a function of inter-click interval. The probabilities were estimated using the logarithmic fit to Fitzpatrick *et al*'s data employed by Carlyon et al (2002), and normalized to that at an inter-click interval of 3.5 ms. This function, like that needed to account for the NH data, decreases in slope over the range shown. If one assumes that the CAP modulation in an alternating-interval pulse train is determined by the ratio of the value of Fitzpatrick *et al*'s recovery function at the two IPIs in that train, then our neural model predicts the pitch matches shown by filled circles without lines in Fig. 3. It can be seen that although this version of the model corresponds to slightly longer periods than the average obtained in experiment 1, it captures the general trends in the data – most notably the tendency for matches to move towards the mean IPI in the stimulus as the overall IPI is increased (Fig. 3b). A *caveat* is that the recovery functions obtained with two-pulse stimuli may not capture the ratio of the CAP amplitudes to pulses occurring after the longer *vs.* shorter inter-pulse intervals in a pulse train. To illustrate this, the dotted line in Fig. 8 shows the amplitude of the GP CAP to the second pulse of a pulse train as a proportion of that to the first, using data obtained in experiment 3b. (These measures should be similar to two-pulse recovery functions provided that the responses to the first two pulses are not strongly influenced by those to the previous pulse trains). Like the function obtained using pulses within a train (solid line), it does not get shallower as IPI is increased from 3.5 to 6.5 ms. It is, however, steeper overall, suggesting that IPI has a larger effect on CAP amplitude at the start of a pulse train than during it.

An alternative explanation is that pitch is affected by some additional source of refractoriness, over and above that observed in the auditory nerve. This could arise either from the output of the AN passing through a second neural stage prior to the “more central”

²The choice of a log function was motivated by other data in the literature (e.g. Fitzpatrick et al., 1999), but a fit using a compressive power function yielded similar results.

neurons, or to refractoriness inherent to those more central neurons. To quantify the additional refractoriness needed, we divided the function fit to the NH data (squares) from the measured GP function (triangles), and plotted this ratio by the dot-dashed lines in Fig 9. Note that this function describes the “gain” applied to a pulse after a given IPI, *relative to that at an IPI of 3.5 ms*. As refractoriness should reduce the amplitude of neural responses, we would expect the *absolute* value of the gain to be less than 1. The curve flattens off above about 4 ms, and the best-fitting logarithmic fit is $y = 0.089 \ln(\Delta t) + 0.9$

B. More general models of pitch.

Our measurements have, for reasons described in the Introduction, been restricted to situations where resolved harmonics are absent. However, the general idea that neural refractoriness can influence “purely temporal” pitch perception could be incorporated into more general models of pitch. As an example, we consider the recent model proposed by Wiegand and colleagues (Wiegand and Meddis, 2001; Wiegand and Winter, 2001). They proposed that pitch may be coded by populations of chop-S neurons in the cochlear nucleus, where each population consists of neurons with a given chopping rate (CR) but a range of characteristic frequencies (CFs). They showed, both by computer simulations and reproductions of GP recordings made by others (Wiegand and Meddis, 2001; Winter et al., 2001), that when a population is stimulated by a sound whose F0 is equal to CR, the neurons in that population show an enhanced tendency to fire at a rate equal to CR. Importantly, this enhancement also occurs when the F0 is an integer multiple of CR, reflecting the ability of chop-S units to “skip” input spikes. As a result, chop-S units with CRs equal to the period of a harmonic complex will chop at $CR = 1/F_0$, even when their CFs are tuned to higher (but still resolved) harmonics of that F0. In the absence of refractory effects, we might expect our (unresolved) 4-6 stimulus to produce enhanced temporal firing in populations with CRs equal to the reciprocals of 4, 6, and 10 ms (the latter representing the ability of chop-S neurons to skip input spikes). If, as we have shown, refractory effects cause the AN input to chop-S neurons to be amplitude-modulated, then we might expect this to increase the enhancement in those populations with CRs of 100 Hz (the reciprocal of 10 ms). Pitch might then be judged from a weighted sum of Chop-S populations with CRs that produce the greatest temporal enhancement. We should note that, in fact, Wiegand and Meddis propose a refractory period of only 0.75 ms, so it is unlikely that the current implementation of the peripheral stages of their model could capture the CAP modulation observed here. However, this could easily be modified by changing the refractory period of the model.

An additional issue facing the Wiegand and Meddis model, shared with our own simple account, is that, when NH listeners are allowed to adjust the period of an isochronous stimulus to match that of a 4-6 pulse train, they never produce matches to a period of 10 ms (Carlyon, 2002). Instead, the resulting distribution of matches is unimodal, suggesting that subjects either do not have conscious access to the separate 4, 6, and 10-ms intervals present in the CAP, or always choose to combine them into a summary measure rather than sometimes matching to one or another of these periods. In contrast, McKay and Carlyon's (1999) MDS study showed that, with *physically* amplitude modulated stimuli (Fig. 1b), listeners *were* sensitive to both the modulator and carrier rates. This difference may have been due to differences in the depth of CAP modulation produced by the different stimuli in the two studies, to the fact that in McKay & Carlyon's study the carrier and modulator rates were harmonically related, or to differences in measurement procedure (pitch judgements vs. MDS).

VI. SUMMARY & CONCLUSIONS

An important topic in the study of pitch perception is the relationship between the representation of the stimulus at the level of the AN and the perceived pitch. The

experiments described here compared behavioral measures of “purely temporal” pitch perception with measurement of auditory nerve activity obtained from GPs and humans, using very similar stimuli. The absence of resolved harmonics in our stimuli greatly simplifies this comparison, and allows a quantitative comparison between physiology and behavior. This feature was used to compare pitch judgements to the predictions of a simple model in which pitch is derived from 1st-order intervals in the combined responses of many AN fibers, as measured by the CAP. According to the model, an array of “more-central” neurons, whose thresholds are uniformly distributed, fire whenever the CAP exceeds threshold. Pitch is then estimated from an unweighted average of the 1st-order intervals in the outputs of these more central neurons.

The results show that the CAP to equal-amplitude alternating-interval stimuli is amplitude modulated both in NH humans and GPs, and that this AM is constant over the 50-dB range of levels studied in the GP. The presence of AM can qualitatively account for the finding that the pitch of such stimuli corresponds to a period that is slightly longer than the mean interval present in the stimulus. Importantly, its level-independence is consistent with our behavioral finding that the pitch of “4-6” stimuli is similar to that observed in a previous study using a 24-dB lower level. This helps resolve the potentially conflicting findings that temporal pitch is dominated by 1st-order intervals in the *stimulus* (Kaernbach and Demany, 1998; Plack and White, 2000; Kaernbach and Bering, 2001; Yost et al., 2005), but that codes based on 1st-order statistics of the responses of individual neurons are strongly level-dependent (Cariani and Delgutte, 1996; McKinney and Delgutte, 1999). The resolution occurs because a model based on 1st order intervals in the neural response can produce realistic pitch estimates that are level-independent, provided that the summary statistic is derived *after* the responses of many neurons have been combined. There are, however, quantitative discrepancies between the predictions of the model and the variation in pitch between 3.5-5.5, 4-6, and 4.5-6.5 stimuli. We discuss this discrepancy in terms of possible species difference and of the effects of refractoriness in neural stages central to the AN.

VII. REFERENCES

- Cariani PA, Delgutte B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 1996; 76:1698–1716. [PubMed: 8890286]
- Carlyon RP. The effects of two temporal cues on pitch judgements. *J. Acoust. Soc. Am.* 1997; 102:1097–1105.
- Carlyon RP. Temporal pitch mechanisms in acoustic and electric hearing. *J. Acoust. Soc. Am.* 2002; 112:621–633. [PubMed: 12186042]
- Carlyon RP, Shamma S. An account of monaural phase sensitivity. *J. Acoust. Soc. Am.* 2003; 114:333–348. [PubMed: 12880045]
- Carlyon RP, van Wieringen A, Long CJ, Deeks JM, Wouters J. Temporal pitch mechanisms in acoustic and electric hearing. *J. Acoust. Soc. Am.* 2002; 112:621–633. [PubMed: 12186042]
- Eggermont JJ, Spoor A. Cochlear adaptation in guinea pigs: a quantitative description. *Audiology.* 1973; 12:193–220. [PubMed: 4722021]
- Fitzpatrick DC, Kuwada S, Kim DO, Parham K, Batra R. Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex. *J. Acoust. Soc. Am.* 1999; 106:3460–3472. [PubMed: 10615686]
- Kaernbach C, Bering C. Exploring the temporal mechanisms involved in the pitch of unresolved harmonics. *J. Acoust. Soc. Am.* 2001; 110:1039–1047. [PubMed: 11519572]
- Kaernbach C, Demany L. Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *J. Acoust. Soc. Am.* 1998; 104:2298–2306. [PubMed: 10491694]
- Kim DO, Molnar CE, Matthews JW. Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure. *J. Acoust. Soc. Am.* 1980; 67:1704–1721. [PubMed: 7372925]

- Levitt H. Transformed up-down methods in psychophysics. *J. Acoust. Soc. Am.* 1971; 49:467–477. [PubMed: 5541744]
- Loeb GE. Are cochlear implant patients suffering from perceptual dissonance? *Ear and Hearing.* 2005; 26:435–450. [PubMed: 16230894]
- Loeb GE, White MW, Merzenich MM. Spatial cross-correlation. *Biol. Cybernetics.* 1983; 47:149–163. [PubMed: 6615914]
- Mahendran S, Bleeck S, Winter IM, Baguley DM, Axon PR, Carlyon RP. Human auditory nerve compound action potentials and long latency responses. *Acta Oto-Laryngologica.* 2007 in press.
- McDermott HJ. Music perception with cochlear implants: a review. *Trends in Amplification.* 1997; 8:49–82. [PubMed: 15497033]
- McKay CM, Carlyon RP. Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains. *J. Acoust. Soc. Am.* 1999; 105:347–357. [PubMed: 9921661]
- McKay CM, McDermott HJ. Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals. *J. Acoust. Soc. Am.* 1998; 104:1061–1074. [PubMed: 9714925]
- McKay CM, McDermott HJ, Clark GM. Pitch matching of amplitude modulated current pulse trains by cochlear implantees: the effect of modulation depth. *J. Acoust. Soc. Am.* 1995; 97:1777–1785. [PubMed: 7699159]
- McKinney MF, Delgutte B. A possible neurophysiological basis of the octave enlargement effect. *J. Acoust. Soc. Am.* 1999; 106:2679–2692. [PubMed: 10573885]
- Moore, BCJ. *An Introduction to the Psychology of Hearing.* New York: Academic; 1989.
- Moore, BCJ.; Carlyon, RP. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Plack, CJ.; Oxenham, AJ., editors. *Springer Handbook of Auditory Research: Pitch Perception.* Springer-Verlag; 2005. p. 234–277.
- Moore BCJ, Glasberg BR, Flanagan HJ. Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure. *Journal of the Acoustical Society of America.* 2006; 119:480–490. [PubMed: 16454302]
- Moore BCJ, Glasberg BR, Peters RW. Relative dominance of individual partials in determining the pitch of complex tones. *J. Acoust. Soc. Am.* 1985; 77:1853–1860.
- Murnane OD, Prieve BA, Relkin EM. Recovery of the human compound action potential following prior stimulation. *Hearing Research.* 1998; 124:182–189. [PubMed: 9822915]
- Oxenham AJ, Bernstein JGW, Penagos H. Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences of the United States of America.* 2004; 101:1421–1425. [PubMed: 14718671]
- Plack CJ, White LJ. Pitch matches between unresolved complex tones differing by a single interpulse interval. *J. Acoust. Soc. Am.* 2000; 108:696–705. [PubMed: 10955636]
- Plomp R. Pitch of complex tones. *J. Acoust. Soc. Am.* 1967; 41:1526–1533. [PubMed: 6075560]
- Pressnitzer D, de Cheveigne A, Winter IM. Perceptual pitch shift for sounds with similar waveform autocorrelation. *Acoustics Research Letters Online.* 2001; 3:1–6. <http://ojps.aip.org/ARLO/top.html> Last viewed online 7 August 2007.
- Pressnitzer D, de Cheveigne A, Winter IM. Physiological correlates of the perceptual pitch shift for sounds with similar waveform autocorrelation. *Acoustics Research Letters Online.* 2004; 5:1–6. <http://ojps.aip.org/ARLO/top.html> Last viewed online 7 August 2007.
- Shamma S. Speech Processing in the Auditory System: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.* 1985; 78:1622–1632. [PubMed: 3840813]
- van Wieringen A, Carlyon RP, Long CJ, Wouters J. Pitch of amplitude-modulated irregular-rate stimuli in electric and acoustic hearing. *J. Acoust. Soc. Am.* 2003; 114:1516–1528. [PubMed: 14514205]
- Wiegand L, Meddis R. The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus. *Journal of the Acoustical Society of America.* 2001; 115:1207–1218. [PubMed: 15058342]

- Wiegrebe L, Winter IM. Temporal representation of iterated rippled noise as a function of delay and sound level in the ventral cochlear nucleus. *J. Neurophysiol.* 2001; 85:1206–1219. [PubMed: 11247990]
- Winter I, Wiegrebe L, Patterson RD. The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea pig. *J. Physiol. (London)*. 2001; 537:553–566. [PubMed: 11731585]
- Yost WA, Mapes-Riordan D, Shofner W, Dye R, Sheft S. Pitch strength of regular-interval click trains with different length “runs” of regular intervals. *Journal of the Acoustical Society of America*. 2005; 117:3054–3068. [PubMed: 15957774]

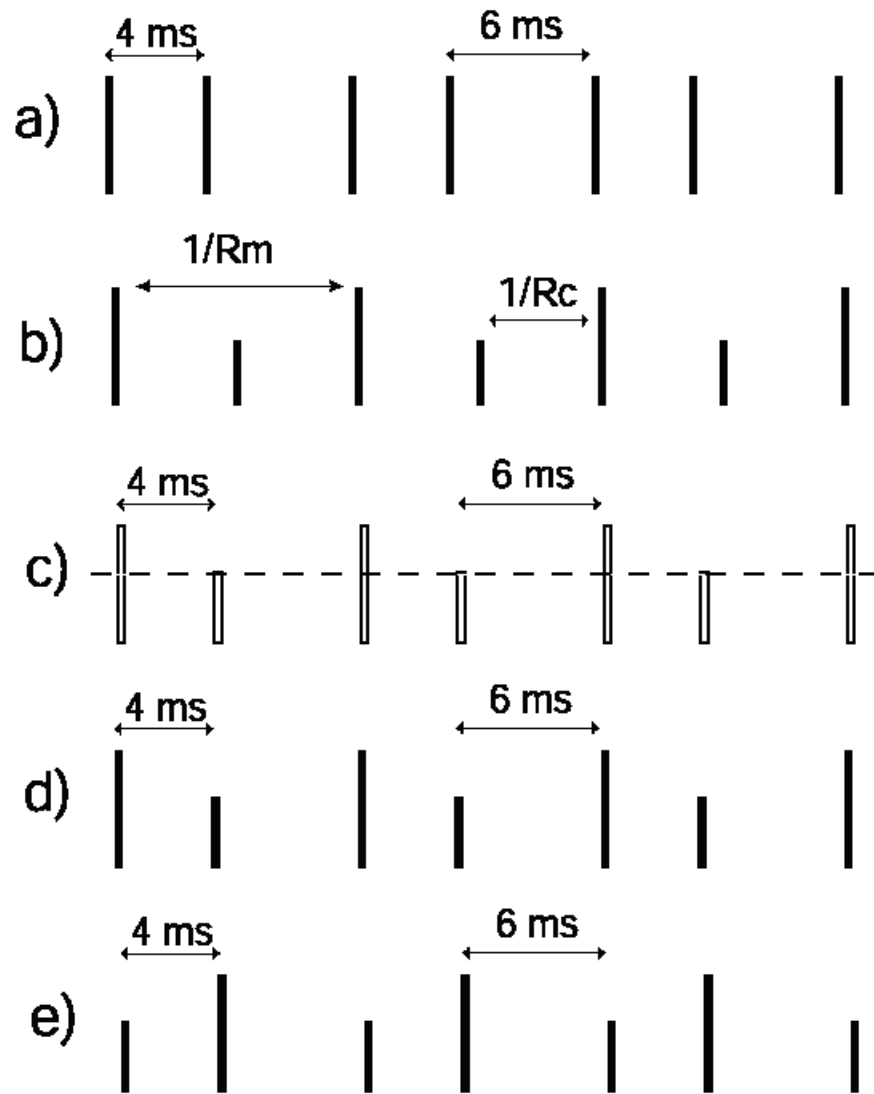


Fig 1.
Solid bars show schematic illustrations of some of the stimuli used in this and other studies. Only the first seven pulses in each train are shown. The open bars in part c) illustrate a possible pattern of CAP responses. Further details are given in the text.

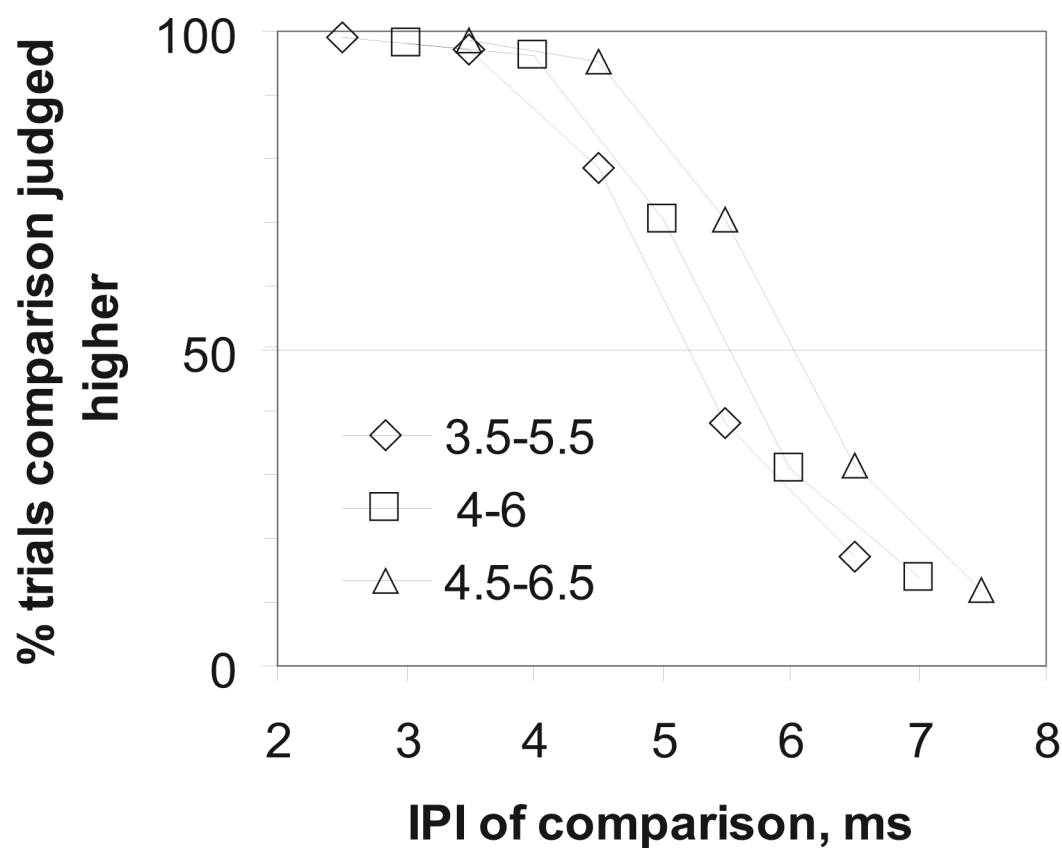
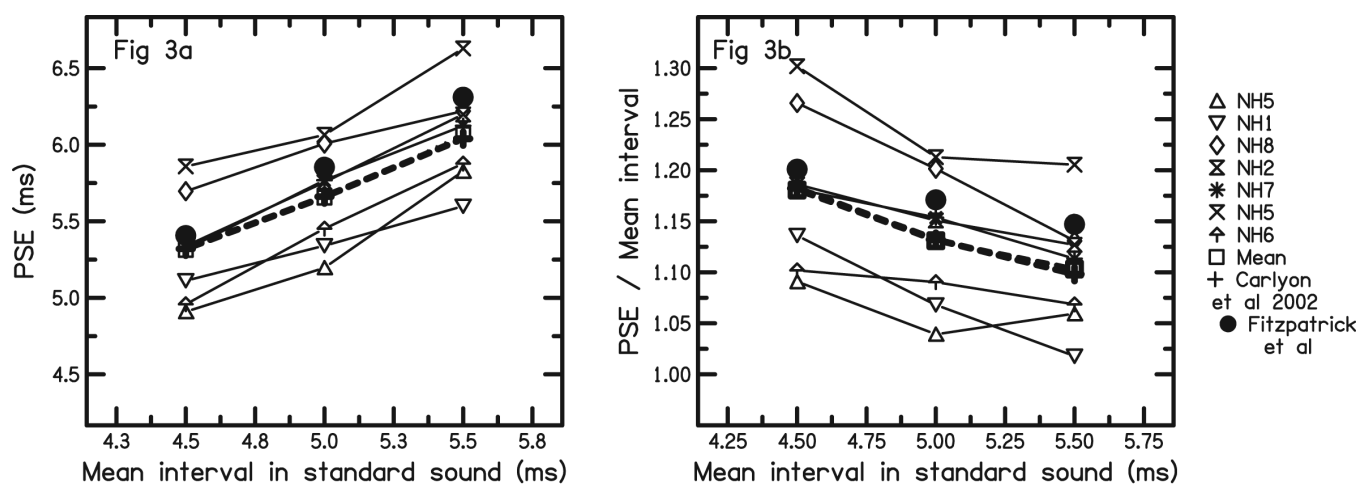


Fig 2. Psychometric functions showing the percentage of trials in which the isochronous comparison sound, whose period is given on the abscissa, was judged higher than 3.5-5.5 (diamonds), 4-6 (squares), and 4.5-6.5 (triangles) standard stimuli. Data are averaged across the NH listeners of experiment 1.

**Fig 3.**

Part a) shows the Point of Subjective Equality (“PSE”) derived from the psychometric functions of experiment 1, for 7 NH listeners. The ordinate shows the mean interval in each of the three standard sounds tested. Mean data are shown by the heavy dashed line joining squares. The prediction of Carlyon et al (2002)’s model is shown by the heavy dashed line joining “plus” signs. These two heavy curves overlap, testifying to the success of the model. Part b) shows the same data, with the PSEs divided by the mean interval in each standard. In both parts of the fig., predictions based on the recovery function described by Fitzpatrick et al (1999) are shown by filled circles.

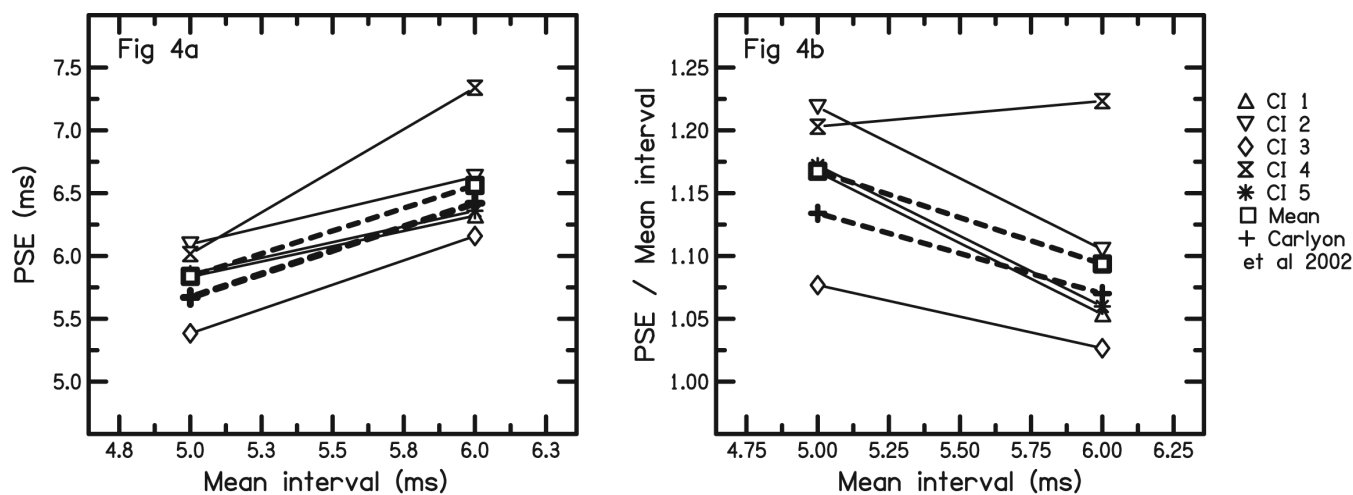
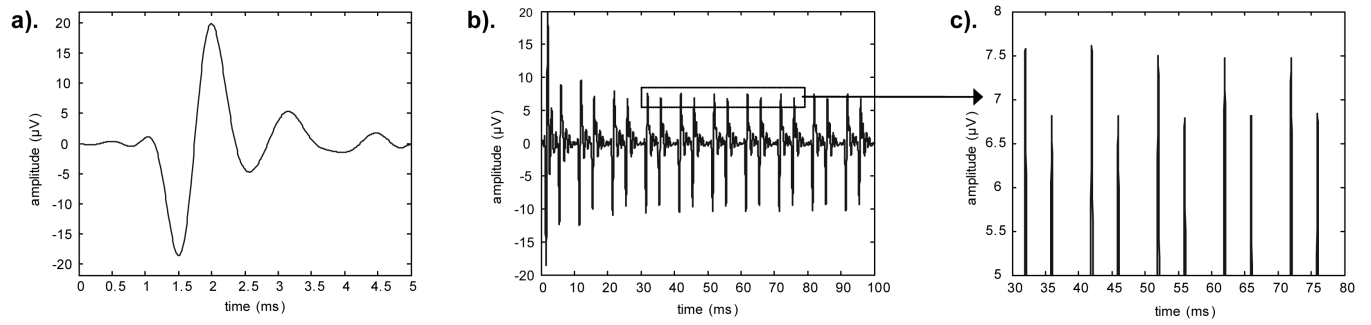
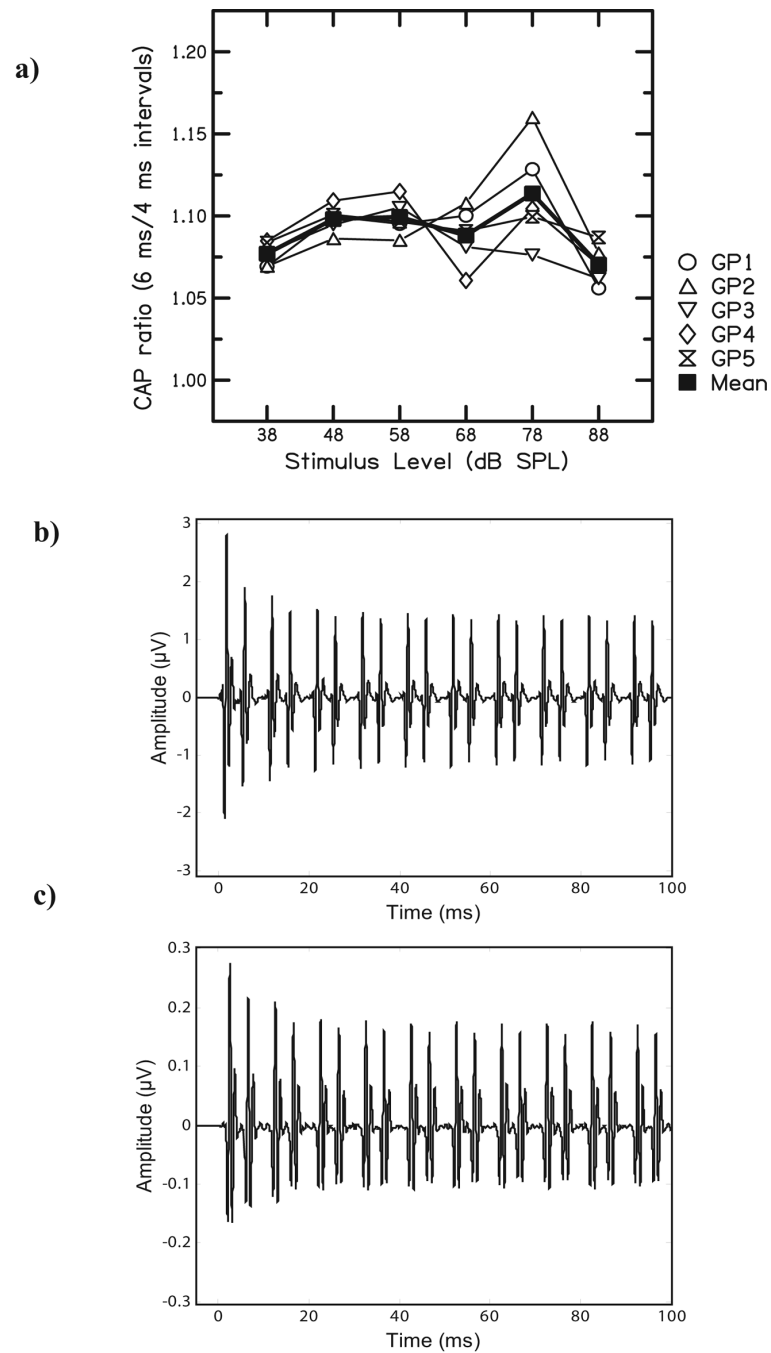


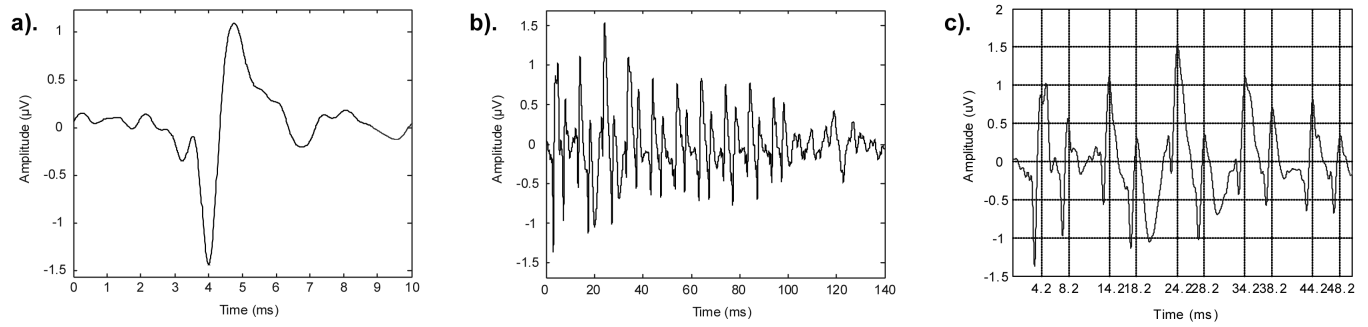
Fig 4.
As Fig. 3, except for the five CI listeners of experiment 2.

**Fig 5.**

Part a) shows the CAP to a single pulse from one GP of experiment 3a. Part b) shows the response to a 78-dB-SPL 4-6 pulse train in the same animal. The area shown by the dashed box is expanded and illustrated in part c).

**Fig 6.**

Part a): Lines connecting symbols show the ratio of CAP amplitudes after 6- vs. 4-ms intervals in a 4-6 pulse train, for each GP of experiment 3a, as a function of stimulus level. The heavy line without symbols shows the mean data. Parts b) and c) illustrate the level-independence of the AN response by plotting the CAP waveform for GP2 at levels of 88 and 38 dB SPL, respectively.

**Fig 7.**

Part a) shows the CAP to a stimulus consisting of the first pulse of a 4-6 pulse train, in listener NH2. Part b) shows the response to part of a 4-6 pulse train in listener NH7. Part c) shows a zoomed-in portion of part b). The vertical gridlines are spaced, alternately, by 4 and 6 ms.

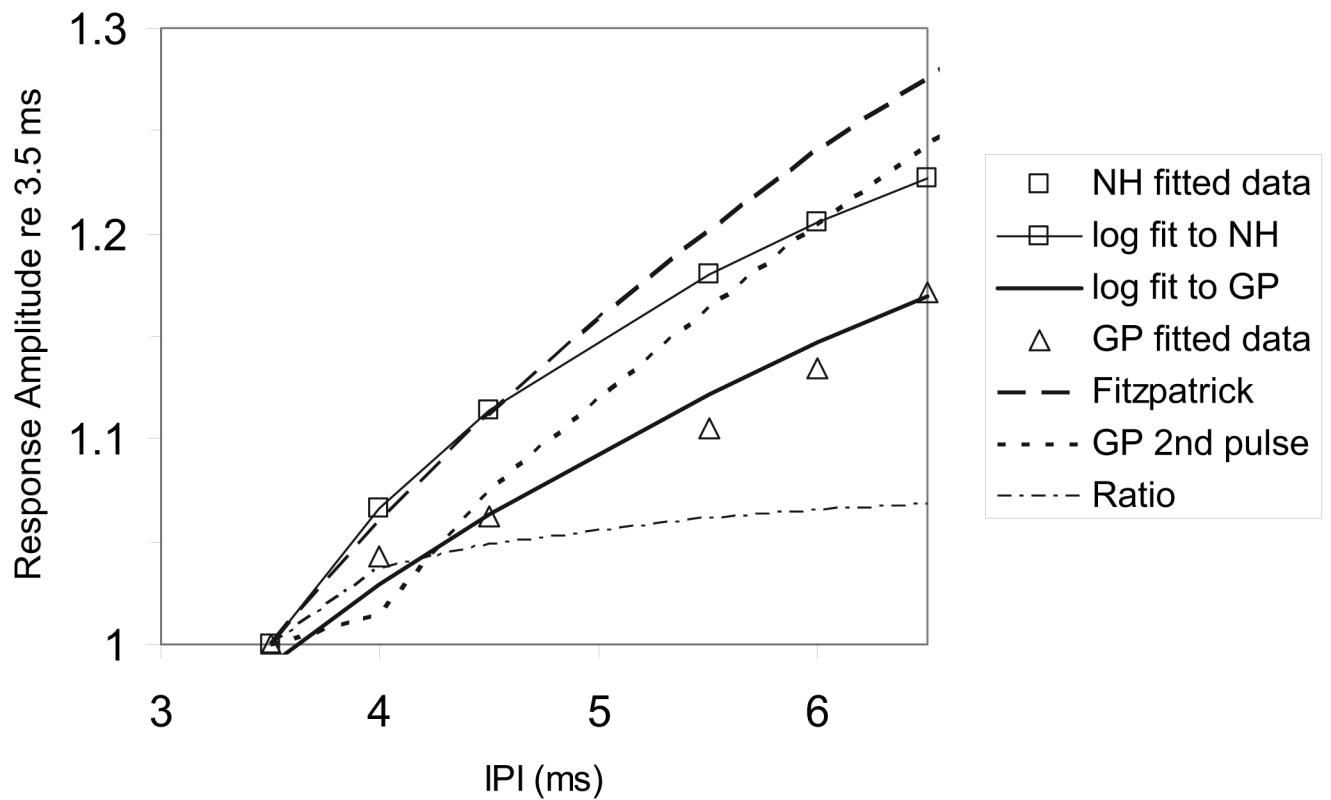


Fig 8.

The squares show the recovery function, expressed as response amplitude *re* that at IPI=3.5 ms, necessary for the neural model to account for the NH pitch data from experiment 1. The solid line passing through these points represents the best fit to these data using a logarithmic function (see text for details). The triangles show the points derived from the GP data of experiment 3b: the ratio between the amplitudes at 5.5 vs 3.5, 6 vs 4, and 6.5 vs 4.5 ms reflect the depth of AM in the CAP response to the 3.5-5.5, 4-6, and 4.5-6.5 stimuli respectively. The vertical distances between other delays (e.g. 3.5 vs 4 ms) were adjusted to provide the best logarithmic fit to the data, shown by the bold solid line. The ratio between these first two curves (faint and bold solid lines) is shown by the dot-dashed line. The bold dashed line shows the two-pulse recovery functions for the cat AN described by Fitzpatrick et al (1999), as fit by Carlyon et al (2002). The dotted line shows the CAP amplitude to the second pulse in each train as a proportion of that to the first, using data obtained from experiment 3b.

Table 1

Details of the cochlear implant patients who took part in experiment 2. CSOM refers to chronic serous otitis media. NI refers to noise-induced hearing loss.

Subject	Age (years)	Etiology	Age of Onset	Implant date	Device
CI 1	58	Familial	47 years	Nov. 1999	CI 24M
CI 2	68	Meniere's/CSOM	Progressive	June 1996	CI 22
CI 3	38	Labyrinthitis	15 years	Nov. 1995	CI 22
CI 4	64	Idiopathic	18 years	April 1996	CI 22
CI 5	77	Otosclerosis / NI	22 years	March 2001	CI 24M

Table 2

a) Percentage difference between the amplitudes of CAPs measured after the shorter vs. longer intervals for the unmodulated stimuli of experiment 3b. Data from 3 GPs are shown, and are averaged across conditions where the stimulus started with the shorter of the two possible intervals (e.g. 4 ms in the “4-6” pulse train) and where it started with longer interval (e.g. 6 ms). An exception occurred for GP7 in the 8-12 condition, where, due to an error, only stimuli starting with the 8-ms interval were used.

b) is similar to a) but for the 4-6 stimuli of experiment 3b in which the pulses occurring after the longer or shorter intervals could be attenuated by 2 or 6 dB, thereby producing amplitude modulation. When modulated, only stimuli starting with the shorter of the two intervals were used, and for consistency analysis of the unmodulated stimuli was restricted to those starting with the shorter interval. (This is why the data for the unmodulated stimuli can differ slightly from those shown for the same stimuli in part a). c) is similar to b) except for the 8-12 stimuli.

a)		Stimulus			
Animal	3.5-5.5	4-6	4.5-6.5	8-12	
GP6	13.0	10.1	9.4	9.5	
GP7	7.3	6.1	4.6	14.4	
GP8	11.1	10.1	17.0	12.7	
Mean	10.5	8.7	10.3	12.2	
b)		Stimulus			
Animal	4-6L6	4-6L2	4-6	4-6S2	4-6S6
GP6	-50.4	-21.3	9.5	57.9	144.2
GP7	-13.4	-8.3	7.4	24.1	35.4
GP8	-5.1	-2.3	7.7	21.7	25.6
Mean	-23.0	-10.6	8.2	34.6	68.4
c)		Stimulus			
Animal	8-12L6	8-12L2	8-12	8-12S2	8-12S6
GP6	-43.4	-14.9	9.5	44.5	130.8
GP7	9.3	3.3	15.1	26.9	13.8
GP8	4.3	2.6	12.7	23.0	18.2
Mean	-9.9	-3.0	12.4	31.4	54.3