JARO 18: 789–802 (2017)
DOI: 10.1007/s10162-017-0632-x

JARO
Journal of the Association for Research in Otolaryngology

Research Article

Vocoder Simulations Explain Complex Pitch Perception Limitations Experienced by Cochlear Implant Users

Anahita H. Mehta[1] and Andrew J. Oxenham[1]

[1]*Department of Psychology, University of Minnesota, 75 East River Parkway, Minneapolis, MN 55455, USA*

Received: 23 February 2017; Accepted: 26 June 2017; Online publication: 21 July 2017

ABSTRACT

Pitch plays a crucial role in speech and music, but is highly degraded for people with cochlear implants, leading to severe communication challenges in noisy environments. Pitch is determined primarily by the first few spectrally resolved harmonics of a tone. In implants, access to this pitch is limited by poor spectral resolution, due to the limited number of channels and interactions between adjacent channels. Here we used noise-vocoder simulations to explore how many channels, and how little channel interaction, are required to elicit pitch. Results suggest that two to four times the number of channels are needed, along with interactions reduced by an order of magnitude, than available in current devices. These new constraints not only provide insights into the basic mechanisms of pitch coding in normal hearing but also suggest that spectrally based complex pitch is unlikely to be generated in implant users without significant changes in the method or site of stimulation.

Keywords: cochlear implants, pitch, vocoder, melody discrimination

INTRODUCTION

Pitch perception is a critical component of auditory and speech perception (McDermott 2004; Oxenham 2012). It provides cues for linguistic features, such as intonation and emphasis contrasts (Highnam and Morris 1987), as well as paralinguistic features, such as age, gender, and emotional state of the talker (Lieberman and Michaels 1962; Abberton and Fourcin 1978; Titze 1989). Pitch is also intrinsically related to music perception, as it conveys crucial information about the melody, harmony, and tonality of sounds. For harmonic complex tones, such as those produced by musical instruments and the human voice, pitch is determined primarily by the low-numbered harmonics (Plomp 1967; Ritsma 1967), which are thought to be spectrally resolved at the level of the auditory periphery (Moore and Gockel 2011; Oxenham 2012). Hearing-impaired listeners commonly experience pitch processing deficits that may arise in part from reduced spectral resolution (Glasberg and Moore 1986; Arehart 1994; Bernstein and Oxenham 2006). The spectral resolution of cochlear implants (CIs) is limited by the number of electrodes or channels (between 12 and 24 in current devices) and by the interactions of electrical current between adjacent electrodes. Because of this lack of resolution, CI users do not have access to the pitch produced by spectrally resolved harmonics and instead receive much weaker pitch information via the periodic fluctuations of a complex tone's temporal envelope (Kong et al. 2009; Carlyon et al. 2010; Macherey et al. 2011; Fielden et al. 2015; Cosentino et al. 2016). This weaker form of pitch, which is also perceived by normal-hearing listeners when presented only with high-numbered spectrally unresolved harmonics (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994), is generally only available for fundamental frequencies (F0s) below about 300 Hz in CIs (Carlyon et al. 2010), is strongly degraded by room reverberation (Qin and Oxenham 2005; Sayles and Winter 2008), and cannot

Correspondence to: Anahita H. Mehta · Department of Psychology · University of Minnesota · 75 East River Parkway, Minneapolis, MN 55455, USA. Telephone: (612) 626-3258; email: mehta@umn.edu

convey more than one pitch at a time (Carlyon 1996; Micheyl et al. 2010; Oxenham and Kreft 2014).

The inability to perceive strong pitch cues via CIs underlies some of the difficulties experienced by CI users in complex acoustic environments. For instance, CI users have difficulty understanding speech in situations with multiple talkers (Stickney et al. 2004; Rosen et al. 2013), and have difficulty recognizing simple melodies (Kong et al. 2004; McDermott 2004; Galvin et al. 2007; Zeng et al. 2014). Although it is believed that the limited number of independent spectral channels explains the lack of complex spectral pitch perception in CI users, it is not known how many channels would be needed, or how independent the information from each channel would need to be to restore such pitch via CIs. Some studies have addressed this question using vocoder-based simulations of CI stimulation in normal-hearing listeners (Kong et al. 2004; Crew et al. 2012). Their results have suggested that as few as 16 channels may be sufficient to elicit pitch (Kong et al. 2004), even when channel interactions are taken into account (Crew et al. 2012). If true, it would suggest that only an incremental improvement in current CIs would be needed to restore some complex spectral pitch. However, a number of factors make the interpretation of the previous results difficult. First, the parameters used in those studies do not rule out the influence of temporal-envelope cues. Although temporal-envelope cues for pitch perception are weak, they can be sufficient to convey melodic information (Burns and Viemeister 1976). Therefore, it is important to rule them out completely to test specifically for spectral cues. Second, the tasks in the previous studies either involved familiar melody recognition or contour recognition, tasks that do not explicitly require the extraction of the F0 (McDermott et al. 2008), meaning that good performance could be achieved by simply extracting the lowest spectral edge of the stimuli, rather than the F0 itself.

The aim of the current study was to determine the spectral resolution necessary to extract pitch from low-numbered, spectrally resolved harmonics, in terms of number of channels and spectral interactions. Noise-excited envelope vocoders were used to simulate aspects of CI processing (Dudley 1939; Shannon et al. 1995). Varying degrees of channel interaction were simulated by using filters with different spectral slopes (Fu and Nogaki 2005; Bingabr et al. 2008). We ensured that the F0 had to be extracted from low-numbered resolved harmonics by band-pass filtering the stimuli before vocoding, so that the lower spectral edge of the stimuli did not provide usable information for the task. We removed potential temporal-envelope pitch cues by low-pass filtering the temporal envelope at 50 Hz. We also limited the information available in the pitch contour of the stimuli by using a melody discrimination task that required the detection of small changes in pitch, involving a single scale step (one or two semitones) that never resulted in a reversal of the melodic contour (Oxenham et al. 2011). The results show that a minimum of 32 channels without any spectral overlap is required for extracting complex spectral pitch. When spectral overlap between channels is introduced, spectral resolution corresponding to at least 64 channels with 72 dB/octave overlap is needed to provide the minimal spectral information necessary to elicit usable pitch.

GENERAL METHODS

Participants

Participants for all experiments were tested to ensure that they had audiometric hearing thresholds no greater than 15 dB hearing level (HL) at octave frequencies from 250 Hz to 8 kHz. No participant had a history of neurological or hearing damage. Written informed consent was provided by each participant, and all participants were compensated for their time. The experiment was conducted at the University of Minnesota Twin Cities. The University of Minnesota Institutional Review Board gave approval for this experiment.

Vocoder Parameters

For experiments 1–4, the stimuli were vocoded with noise carriers. Although sine wave vocoders have been shown to better represent CI users' performance on pitch-related tasks (Luo et al. 2007), they can also potentially add extraneous pitch cues. The acoustic pure and complex tones were first passed through a filter bank of logarithmically spaced frequency bands using very high-order finite impulse response (FIR) filters ($N > 1000$), which ensured a flat overall spectral response (<0.1 dB ripple in the passband) and essentially no spectral overlap between channels. These were the analysis filters used for all the experiments. The impulse response of the filters was time aligned with a latency of about 20 ms. The frequency range for the entire filter bank (from the lowest to the highest cutoff frequencies) was 200–6000 Hz, and the number of channels was varied based on the experiment. Following band-pass filtering, the temporal envelopes in each channel were computed using a Hilbert transform followed by a fourth-order Butterworth low-pass filter with a cutoff frequency of 50 Hz to remove periodicity cues from the envelope. The resulting envelopes were then used to modulate independent samples of Gaussian white noise, and the resulting stimuli were then re-filtered through the synthesis filter bank. The slopes of the synthesis filter bank were varied according the parameters in the various experiments using either very high-order FIR filters ($N > 1000$) for the no-overlap conditions or Butterworth filters ranging from fourth (24 dB/oct) to 60th order (360 dB/oct), as described in each of the individual experiments as follows. Finally, the output stimulus was generated by summing the outputs across all channels.

Experimental Setup

All stimuli were generated digitally and were converted to analog at a sampling rate of 48 kHz via a Lynx Studio L22 sound card. All testing was carried out in a sound-attenuating booth. Stimulus presentation and response collection were controlled using the AFC software package (Ewert 2013) under MATLAB (MathWorks, Natick, MA). The stimuli were presented diotically at 65 dB sound pressure level (SPL) via HD 650 headphones (Sennheiser, Wedemark, Germany). For all experiments, all the stimuli were presented in background threshold-equalizing noise (TEN) at 55 dB SPL per equivalent rectangular bandwidth (ERB) around 1 kHz Moore et al. (2000).

EXPERIMENT 1: EFFECT OF NUMBER OF SPECTRAL CHANNELS

Rationale

The aim of experiment 1 was to determine the number of spectral channels required for listeners to accurately perceive spectral pitch, as measured in a melody discrimination task. In this experiment, there was no spectral overlap between adjacent channels.

Participants

Thirteen participants (six males, aged 20–29) took part in this experiment. All participants performed two tasks: Task 1, a training and screening task using non-vocoded stimuli, and Task 2, the task of interest, using noise-vocoded stimuli.

Stimuli

The perception of pitch was measured using the paradigm described by Oxenham et al. (2011). This paradigm involves a two-interval, two-alternative forced-choice melody discrimination task, where each interval consists of a four-note melody. On each trial, the four-note melody was generated from an octave of the diatonic (major) scale, where the first note of the scale (tonic) had an F0 of either 250 or 500 Hz. The notes were selected randomly with uniform distribution and replacement, with the restriction that no three consecutive notes could be the same. Either the notes in the second interval melody were unchanged ("same" trial) or the second or third note of the melody was raised or lowered by one scale step ("different" trial) (see Fig. 1). Listeners were asked to indicate whether the two melodies were the same or different. For example, in the schematic shown in Figure 1, each box corresponds to one trial with two melodies. For the two panels on the left, there is no change in the overall pitches of the notes in the melodies so the correct response would be same. For the two example trials in the panels on the right, the second note is shifted downwards by one step so the correct response would be different. Because the change was only one scale step (either one or two semitones), it never resulted in a reversal of the melodic contour. Experiments 1–4 all used the same melody discrimination task. The only differences across experiments involved modifications in the vocoder parameters. When pure-tone pitch perception was tested, the first melody was comprised of pure tones between 250 and 500 Hz, whereas the second melody was transposed up one octave (500 to 1000 Hz), to avoid listeners being able to compare identical stimuli across the two melodies. When complex-tone pitch perception was measured, the first melody was still comprised of pure tones between 250 and 500 Hz, but the second melody was comprised of harmonic complex tones, band-pass filtered between 1000 and 3000 Hz (filter slopes of 48 dB/octave), with F0s ranging from 250 to 500 Hz. The change from pure to complex tones ensured again that listeners were not able to compare identical stimuli across the two melodies. In order to correctly identify a step change, listeners must extract the pitch of each of the tones. Each tone had a total duration of 300 ms, including 10-ms raised-cosine onset and offset ramps. Within a melody, notes were separated by 200-ms gaps, leading to a tone repetition time of 500 ms, or a tempo of 120 beats per second. These stimulus durations are similar to previous studies investigating melody perception in normal-hearing listeners as well as CI users (Pressnitzer et al. 2001; Kong et al. 2004; Oxenham et al. 2011). All the stimuli were vocoded with the noise vocoder described earlier before being presented to the listeners. The number of channels were 8, 16, 32, and 64. The synthesis filter slopes were designed to have no overlap between channels, so the same FIR filters as the analysis filters were used for the synthesis filter in this experiment (see description in the "General Methods").

Procedure

Task 1 involved participant training and screening. This part consisted of two blocks of 10 runs, with each

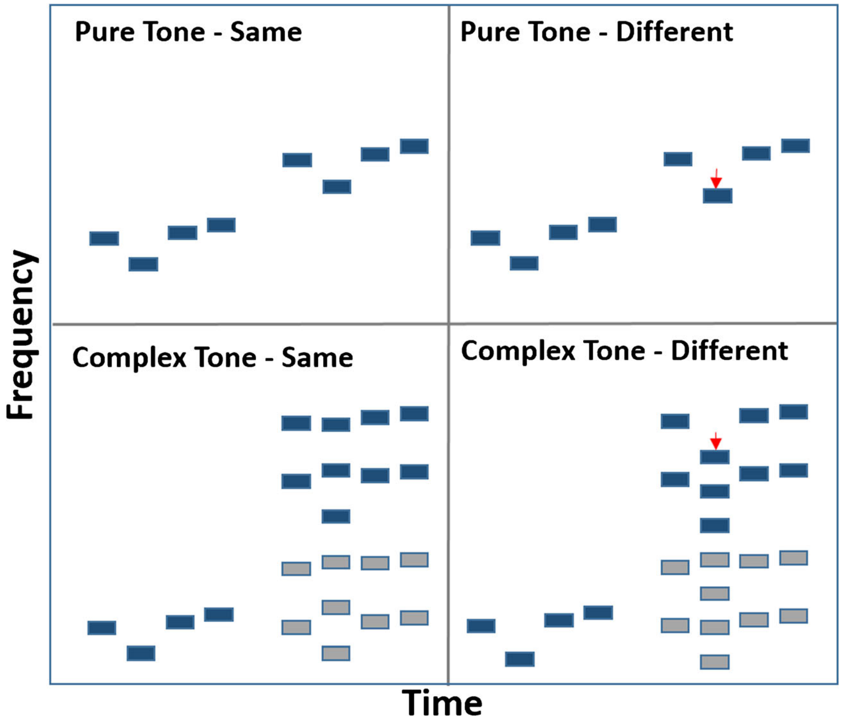

FIG. 1. Schematic representation of the pure and complex tone melody discrimination paradigms used in experiments 1–4. The *top row* shows the pure - tone melody discrimination paradigm, where the second interval always has a melody shifted upwards by an octave. The *bottom row* shows the complex - tone melody discrimination paradigm, where the second interval is made of complex tones whose F0s correspond to the F0s of the tones in the first interval; however, the complex tones are filtered between 1 and 3 kHz so that the F0s are physically not present (the removed components of each tone shown by the *grey blocks*). In both "different" paradigm examples, the second tone is shifted downwards by one scale step (indicated by the *red arrow*).

run consisting of 20 trials. Each block consisted of five runs each for the pure-tone and complex-tone conditions. The training was carried out on the non-vocoded stimuli to ensure that listeners understood and were able to complete the task. Visual feedback (correct/incorrect) was provided after each trial. The participants were required to obtain a score of 80 % correct or higher for three consecutive runs within the two blocks in order to continue with the study. After the training, 10 of the 13 participants successfully passed the screening and were allowed to continue with the study to Task 2. The three participants who failed the screening could still perform the task to an average level of 70 % accuracy, and all had some musical training. Three of the remaining ten participants reported having no formal musical training. Task 2 consisted of five test blocks consisting of ten runs, each corresponding to a different condition, with 20 trials per condition. The ten conditions tested were two types of tones (pure and complex tones) and five vocoder conditions (8, 16, 32, and 64 channels and non-vocoded). No feedback was provided in Task 2 to avoid the possibility that the participants learned to base their judgments on any non-pitch-related cues. The order of presentation of the blocks and runs was randomized.

Behavioral Data Analysis

For each task in experiments 1–4, each participant's performance on the melody perception task was first calculated using d' and then converted to the percentage equivalent proportion correct score assuming unbiased responding (PC_{max}): proportion correct (max for 2AFC) percentage = $(d'/\sqrt{2}) \times 100$ (Macmillan and Creelman 2004). This method gives us a more accurate estimate of sensitivity, independent of bias, as it is derived from the d' scores.

Results

The results of this experiment for both the pure- and complex-tone conditions are shown in Figure 2a. Equivalent proportion correct responses under unbiased responding (PC_{max}) were analyzed for both tasks using a repeated-measures analysis of variance (ANOVA), where the factors were stimulus type (pure or complex tones) and number of channels (8, 16, 32, 64, no vocoder). An overall significant main effect of

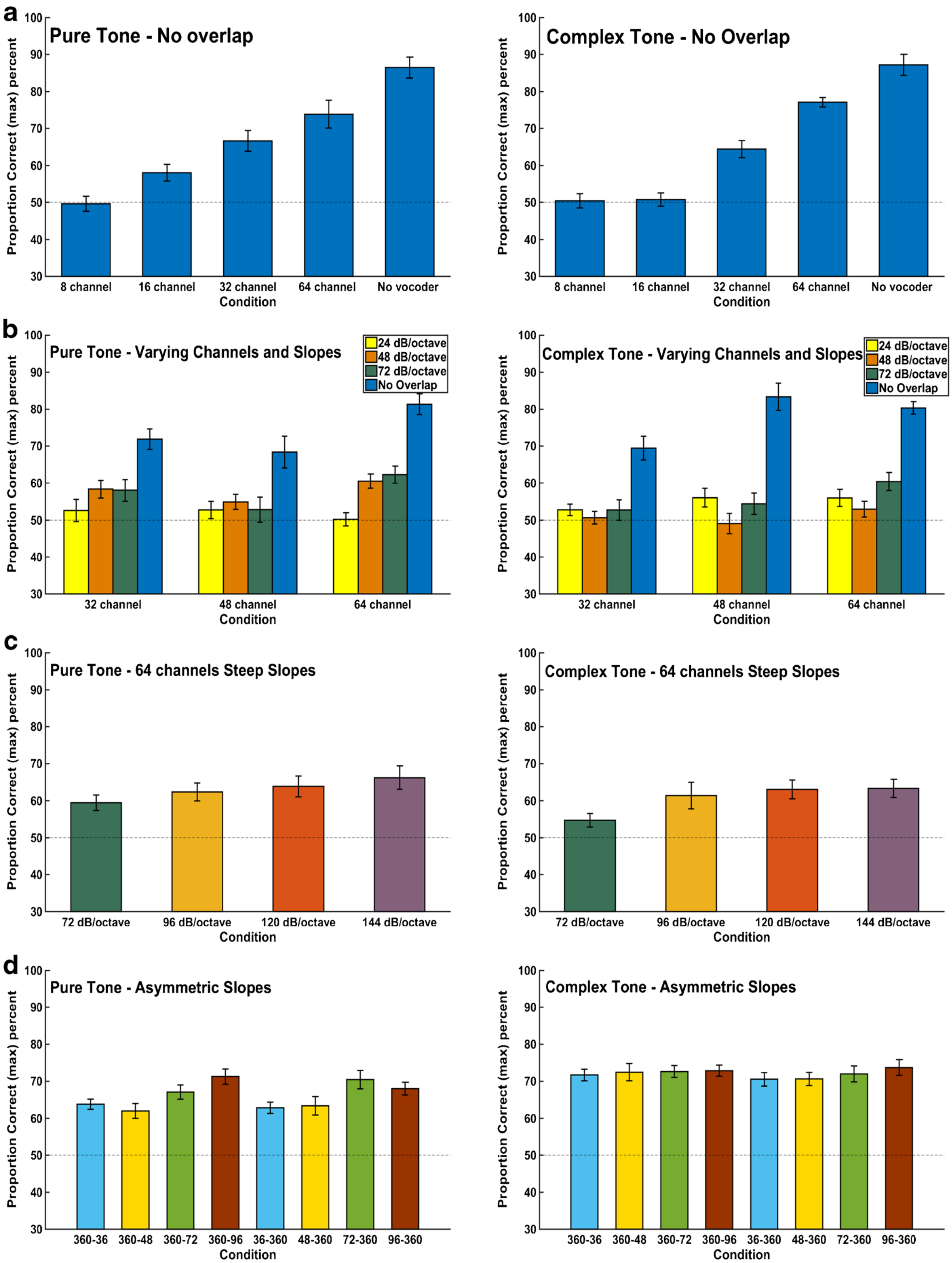

FIG. 2. Average scores in proportion correct (max) for experiments 1–4 for both pure and complex tone conditions. Rows (**a–d**) correspond to experiments 1–4. *Error bars* represent SEM.

the number of channels was found ($F_{4,\ 36} = 78.6$, $P < 0.001$). There was no significant effect of pure vs. complex tones ($F_{1,\ 9} = 0.3$, $P = 0.6$) and no significant interaction between the two factors ($F_{4,\ 36} = 2.4$, $P = 0.07$). Post hoc tests (with Bonferroni adjustment for multiple comparisons; criterion P value = 0.005) indicated that the 32, 64, and no vocoder conditions were significantly different from all other conditions ($P < 0.001$ in all cases); no significant difference was found between 8 and 16 channels ($P > 0.05$). For the pure tones, only performance in the 8-channel condition was not significantly above the 50 % level of chance ($P > 0.05$); performance in all the other conditions exceeded chance ($P < 0.001$ in all cases). For the complex tones, performance was not significantly different from chance in the 8- and 16-channel conditions ($P > 0.05$ in both cases). In order to test for practice effects across blocks, a repeated-measures ANOVA was carried out, again with PC_{max} as the dependent variable, but with factors of tone type (pure or complex), number of channels, and block number. The same results as earlier were observed, along with no significant effect of block number ($F_{4, 36} = 2.1$, $P = 0.1$) and no significant interactions between block number and the other factors. Thus, no evidence for practice effects was observed. Overall, the results suggest that complex pitch perception requires at least 32 channels.

EXPERIMENT 2: INTERACTIONS BETWEEN NUMBER OF CHANNELS AND SPECTRAL SLOPES

Rationale

Experiment 1 indicated that the minimum number of channels required is 32. Although this number is higher than that suggested in some previous studies, even this may be an underestimate when considering CIs, as it does not take into account the effects of interactions between channels. In experiment 2, the same paradigm was used as that in experiment 1, but different filter slopes were used to simulate different degrees of channel interactions.

Participants

Twelve participants (six males, aged 20–29) took part in this experiment. None had previously taken part in experiment 1. All participants had audiometric thresholds of no more than 15 dB HL at octave frequencies from 250 Hz to 8 kHz. No participant had a history of neurological or hearing damage. All participants performed two tasks: Task 1, a training task using non-vocoded stimuli, and Task 2, the main task of interest using noise-vocoded stimuli. Two participants failed to reach the performance criteria for Task 1, leaving 10 participants (four males and six females) to perform Task 2.

Stimuli

The stimulus paradigms for experiment 2 were the same as for experiment 1. All the stimuli were vocoded using the noise vocoder described earlier. To simulate different number of channels, the number of frequency bands was varied (32, 48, and 64 channels). To vary the degree of channel interaction, the slope of the synthesis filters used to create the frequency bands was varied, with slopes of 24, 48, and 72 dB/octave, along with a control no-overlap condition, as in experiment 1.

Procedure

As in experiment 1, Task 1 was a training as well as a screening task, which was carried out using non-vocoded stimuli. The participants were required to obtain a score of at least 80 % correct in three consecutive runs within the two training blocks in order to continue with the study. After the training, ten of the 12 participants successfully passed the screening and were allowed to continue with the study to Task 2.

Task 2 consisted of five test blocks, each consisting of 24 runs. Each run corresponded to a different condition, with 20 trials per condition. The 24 conditions tested were two types of tones (pure vs. complex tones) × three channel conditions (32, 48, and 64 channels) × four spectral slope conditions (24, 48, and 72 dB/octave and no-overlap). Similar to Task 1, the participants were presented in each trial with two melodies of four tones each and were asked to determine if the two melodies were the same or different. Unlike Task 1, participants were not given visual feedback after every trial. The order of presentation of the blocks and runs was randomized for each participant.

Results

The results of this experiment are shown in Figure 2b. The resulting PC_{max} scores were analyzed using a repeated-measures ANOVA with factors of stimulus type (pure or complex tones), number of channels (32, 48, 64), and filter slope (24, 48, and 72 dB/octave and no overlap). A significant main effect of slope was found ($F_{3, 27} = 38.4$, $P < 0.001$), with no other main effects and no significant interactions ($P > 0.05$ in all cases). Post hoc tests (with Bonferroni adjustment for multiple comparisons; criterion P value is 0.008) indicated that the no-overlap conditions were significantly different from all other slope conditions ($P < 0.001$), while the other three conditions with overlap were not significantly different from each other ($P > 0.05$). Furthermore, only the no-overlap conditions produced performance that was significantly above chance for all three channel conditions ($P < 0.005$), whereas of the conditions with shallower filter slopes, only the 72 dB/octave conditions produced above - chance performance, and then only with 64 channels ($P < 0.001$).

EXPERIMENT 3: SPECTRAL SLOPES NEEDED FOR ABOVE - CHANCE PERFORMANCE

Rationale

The results of experiment 2 were unexpected, as we found that even with slopes of 72 dB/octave, performance was generally not significantly above chance. The aim of experiment 3 was to determine what steepness of slopes would be needed, given a large number of channels (64), to elicit accurate complex spectral pitch.

Methods

The 10 participants (four males, aged 20–29) who completed experiment 2 also took part in experiment 3. The same melody discrimination paradigm was used as in experiments 1 and 2. Only the 64-channel vocoder was used in this experiment, together with filter slopes of 72, 96, 120, and 144 dB/oct. As all the participants in experiment 3 had already passed the training/screening task of experiment 2, they only completed the test task here. The task consisted of five test blocks each consisting of eight runs corresponding to different conditions, with 20 trials per run. The eight conditions tested included two types of tones (pure vs. complex tones) with four spectral slope conditions (72, 96, 120, and 144 dB/oct). The task was the same as the previous two experiments. Participants were not given feedback. The order of presentation of the blocks and runs was randomized for each participant.

Results

The results are shown in Figure 2c. The PC_{max} values were submitted to a repeated-measures ANOVA, where the factors were stimulus type (pure and complex tones) and amount of overlap or slope (72, 96, 120, and 144 dB/octave, and no overlap). A significant main effect of slope was found ($F_{3, 27} = 8.86$, $P < 0.001$), but no main effect of stimulus type ($F_{1, 9} = 5.01$, $P = 0.06$) and no significant interaction ($F_{3, 27} = 0.69$, $P = 0.56$). Contrast analysis revealed a linear trend for filter slope ($F_{1, 9} = 21.8$, $P = 0.001$), suggesting a gradual improvement in performance with increasing filter slope. This impression is supported by post hoc tests (with Bonferroni adjustment for multiple comparisons; criterion P value is 0.008), which indicated that only the shallowest (72 dB/oct) and steepest (144 dB/oct) conditions were significantly different from each other ($P < 0.005$). All the conditions produced a performance that was significantly above chance ($P < 0.001$ in all cases). However, even performance with the 144 dB/oct slopes was substantially poorer on average (~65 %) than that found with the steep filters used in the no-overlap conditions with 64 channels (~75 %).

EXPERIMENT 4: EFFECTS OF ASYMMETRIC FILTER SLOPES

Rationale

The results from experiments 1–3 show that very steep filter slopes are critical for eliciting pitch from the vocoder noise bands. The aim of this experiment was to determine if one steep edge would improve performance and, if so, whether the upper or lower filter slope was more important in defining the pitch strength of the stimulus. On one hand, spectral edge pitch (Kohlrausch and Houtsma 1992) has been found to be stronger for the lower-frequency edge than the higher, leading to the prediction that the lower filter slope may limit perception. On the other hand, it may be that any sharp spectral edge is sufficient to induce a pitch percept; in that case, no perceptual asymmetry may be observed between the two slopes. These predictions were tested by measuring melody discrimination using filters that had one steep slope (360 dB/oct) and one slope that decayed at 36, 48, 72, or 96 dB/oct.

Methods

The ten participants (four males, aged 20–29) who took part in experiments 2 and 3 also took part in experiment 4. The experimental paradigm for experiment 4 was the same as for the earlier experiments. All the stimuli were vocoded using the noise vocoder described earlier. In this experiment, only 64 channels were used. One filter slope was always fixed at 360 dB/oct, while the other slope was 36, 48, 72, or 96 dB/oct.

As all the participants in experiment 4 had passed the training/screening task in experiment 2, they only carried out the test task. The task consisted of five test blocks consisting of 16 runs, each corresponding to a different condition, with 20 trials per condition. The 16 conditions tested were two types of tones (pure vs. complex tones) × two sides for steep edge (low- vs. high-frequency edge) × four spectral slope conditions (36, 48, 72, and 96 dB/octave). The task was the same as the previous three experiments. Participants were not given feedback. The order of presentation of the blocks and runs was randomized for each participant.

Results

The results from this experiment are shown in Figure 2d. The PC_{max} values were submitted to a repeated-measures ANOVA where the factors were stimulus type (pure and complex tone), side of steep edge (low and high), and slope of the shallow edge (36, 48, 72, and 96 dB/octave). Significant main effects of stimulus type ($F_{1, 9}$ = 34.3, P < 0.001) and slope of the shallow edge ($F_{3, 27}$ = 7.28, P < 0.001) were observed, along with a significant interaction between those two factors ($F_{3, 27}$ = 3.62, P = 0.024). However, no significant effect of side of steep edge was observed ($F_{1, 9}$ = 0.12, P = 0.73), and no significant interactions with the side of the steep edge were found (P > 0.05 in both cases). Post hoc tests (with Bonferroni adjustment for multiple comparisons; criterion P value is 0.003) indicated that the 72 and 96 dB/octave conditions were significantly different from the 36 and 48 dB/octave conditions (P < 0.001). Contrast analysis revealed a linear trend for filter slope ($F_{1, 9}$ = 24.8, P = 0.001), suggesting a gradual improvement in performance with increasing filter slope. All the conditions produced a performance that was significantly above chance (P < 0.001 in all cases). The results suggest that only one steep filter slope is necessary for accurate pitch perception, and that steep low and high slopes produce equally salient pitches.

EXPERIMENT 5: PITCH MATCHES AND PITCH COMPARISONS

Rationale

Experiments 1–4 investigated the minimum spectral resolution required to extract complex spectral pitch. The results demonstrate that a large number of channels (32 or more) and very steep slopes on at least one side of the spectrum of each channel (greater than 72 dB/oct) are required to induce an accurate pitch sensation. However, it remains unclear what pitch the vocoded stimuli induces, whether it corresponds to the pitch computed from the peaks within the spectrum which would correspond to the F0 of the tone, some transformation of the spectral centroids of the vocoder peaks, or a transformation of the steepest slopes from the vocoder peaks. Here, we used pitch-matching and pitch-comparison paradigms to determine the pitch produced by the vocoded pure and complex tones used in experiments 1–4 and to relate the pitch to the physical characteristics of the vocoded stimuli.

Participants

Fourteen normal-hearing participants (four males, ages 18–29) were recruited to perform a pitch-matching task. All participants had hearing thresholds of less than 15 dB HL at octave frequencies from 250 Hz to 8 kHz. All participants had previous musical training. Eight of the participants had taken part in experiments 2–4; six were newly recruited to the study.

Stimuli and Procedure

Pitch-Matching Task. A pitch-matching paradigm was carried out for both pure tones and complex tones. In both tone conditions, each trial began with a 300-ms vocoded tone followed by a 500-ms silence, and then, a non-vocoded tone. For the pure tone conditions, the trial started with a vocoded pure tone followed by a non-vocoded pure tone. Similarly, for the complex tone condition, the trial started with a vocoded complex tone filtered between 1 and 3 kHz (the same as the complex vocoded tones used in experiments 1–4) followed by a non-vocoded-complex tone, also filtered between 1 and 3 kHz. In both conditions, the listeners had to vary the pitch of the non-vocoded tone until it matched the pitch of the vocoded tone. The tones were embedded in a broadband (50 Hz to 22 kHz) threshold-equalizing noise (TEN; Moore et al. 2000); background noise began 300 ms before the standard tone, continued through the trial, and ended 300 ms after the end of the comparison tone. The tones were vocoded using the same vocoder as the previous experiment. Six stimulus conditions were used: three non-overlapping conditions similar to experiment 2 (32, 48, and 64 channels), two asymmetric slope conditions with 64 channels where either the low and the high pass slopes were 360 and 36 dB/octave or vice versa, and a final symmetric steep slope condition with 64 channels and 72 dB/octave slopes. The F0s of the tones were either 280, 375, or 500 Hz.

The starting frequency or F0 of the variable non-vocoded tone was randomly selected on each block from a uniform distribution on a discrete semitone scale ±18 semitones around the frequency or F0 of the vocoded tone. After each trial, participants could adjust the frequency or F0 of the non-vocoded tone up or down by four, one, or 0.25 semitones; could elect to hear the trial at the same frequencies again; or could indicate that they were satisfied with the pitch match by using virtual buttons on a graphical user interface. Participants were encouraged to bracket the pitch of the reference vocoded tone by adjusting the pitch of the non-vocoded tone below and above that of the reference before making a final decision.

The task consisted of eight blocks, four each for the complex- and pure-tone conditions. Within each run, there were 36 trials. The experiment took place over two to three sessions of 2 h each. This task was not timed, and no feedback was given; participants were encouraged to take their time to ensure accuracy. Due to the challenging nature of this task, an exclusion criterion for the experiment was set. A participant was required to achieve 30 % matching accuracy in pure and complex tones in any condition. Based on this criterion, four participants were excluded from the study, leaving a total of 10 participants (seven females and three males, ages 18–29). Seven of the ten participants in the final dataset had also completed experiment 4.

Pitch Comparison Task. The stimuli used were the vocoded pure and vocoded complex tones used in experiment 5 for the three conditions with overlap (360–36, 36–360, and 72–72 dB/octave). In both tone conditions (pure and complex), each trial began with a 300-ms vocoded tone from one condition followed by a 500-ms silence, and then a vocoded tone from another condition with the same frequency or F0. For example, a trial could have the first tone as a 500-Hz complex tone vocoded with 360–36 dB/octave asymmetric slopes followed by a 500-Hz complex tone vocoded with 36–360 dB/octave asymmetric slopes. The listeners would have to indicate which of the two tones sounded higher.

Listeners were instructed to pick the "higher tone" in a two-interval forced-choice paradigm. A total of 18 runs were presented to the listeners, with each run consisting of 20 trials. Each run corresponded to a different tone (F0s = 280, 375, and 500 Hz) and condition comparison. No feedback was provided. The order of presentation of the blocks and runs was randomized for each participant.

Results

Figure 3 shows the histograms, pooled across all participants and all trials, for both pure and complex tones across all vocoder conditions, with the *x*-axis representing the deviation in semitones from the frequency or F0 of the target tone prior to vocoding. For the analysis of accuracy, the proportion of matches within +/−0.25 semitones of the comparison tone was analyzed using a repeated-measures ANOVA, where the factors were stimulus type (pure or complex tone) and vocoder condition (three non-overlap conditions, two asymmetric conditions, and one symmetric condition). A significant main effect of vocoder condition was observed ($F_{5,\ 45} = 49.6$, $P < 0.001$), with no significant effect of stimulus type ($F_{1,\ 9} = 2.44$, $P = 0.15$) and no significant interaction ($F_{5,\ 45} = 2.25$, $P = 0.065$). Post hoc tests (with Bonferroni adjustment for multiple comparisons; criterion P value is 0.003) revealed that the 36–360 dB/oct asymmetric condition and the 72 dB/octave symmetric condition were significantly different from the other four conditions ($P < 0.001$ in both cases), reflecting the generally lower performance in both conditions. Additionally, the 64 channel no-overlap condition was significantly different from all other vocoder conditions ($P < 0.001$ in all cases), reflecting the fact that performance was the best in this condition. In summary, according to the pitch matches, the modal pitch of the vocoded tones corresponded to spectral peaks of the vocoded stimuli and was roughly the same for the symmetrically and asymmetrically filtered stimuli, although the pitch accuracy was poorer for the asymmetrically filtered stimuli with the shallow slope on the low side. Thus, according to the outcomes of the pitch-matching experiments, the pitch of the vocoded stimulus seems to be determined by the peak frequency of the vocoder filter, and not by the center frequency of the slope, or the spectral centroid of the filter.

In the pitch comparison task, participants were asked to directly compare the pitch produced by the two asymmetric conditions with 360 and 36 dB/oct slopes on either side. All participants responded at a rate of 100 % for both pure and complex tones that the pitch of tones vocoded with asymmetric slopes of 360–36 dB/oct was higher than the pitch of tones vocoded with either asymmetric slopes of 36–360 dB/oct or symmetrical slopes of 72 dB/oct. In addition, all listeners judged at a rate of 100 % that the pitch of pure and complex tones vocoded with symmetrical slopes of 72 dB/oct was higher than the pitch of the tones vocoded with asymmetrical slopes of 36–360 dB/oct. Thus, although the pitch matches revealed no difference in the modal pitch of the two asymmetric conditions (suggesting that the peak of the vocoder filters determines the pitch), direct comparisons resulted in pitch judgments consistent with the spectral centroid of the stimuli, in line with expectations based on the timbral property of brightness (McDermott et al. 2008; Allen and Oxenham 2014).

GENERAL DISCUSSION

Spectral Resolution Required for Complex Spectral Pitch Is Higher than Expected

The results from experiment 1 provide an estimate of the minimum number of spectral channels that is needed to elicit complex spectral pitch. The outcomes suggest that a minimum of 32 channels are required, which with our logarithmically scaled vocoder is roughly equivalent to channel bandwidths of

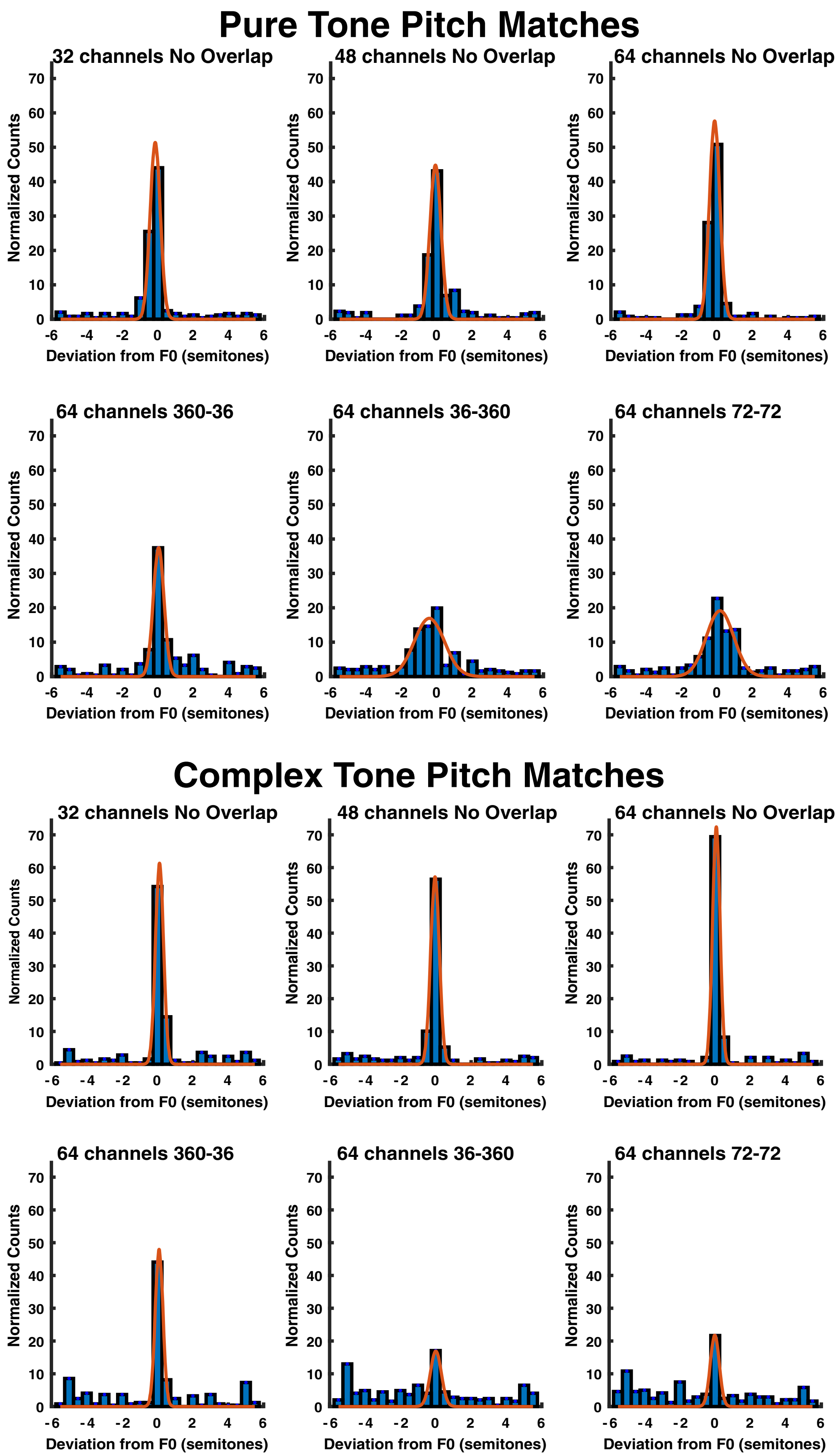

FIG. 3. Average pitch-matching scores for both pure and complex tone conditions for experiment 5. Each histogram indicates the percentage average normalized pitch-matching scores for each vocoder condition fitted with a Gaussian curve.

two semitones or 12 %. In terms of distance along the human cochlea, using Greenwood's (1990) map, this corresponds to roughly 0.6 mm per channel in the central part of the cochlea. Note that the estimated number of channels can be considered a lower bound, because it does not take into account any potential effects of spectral spread or non-uniform neural survival.

Experiments 2 and 3 show that even with 32 or 64 channels, pitch perception is very limited when some spectral spread is incorporated in the simulations: even filter slopes as steep as 72 dB/oct were not sufficient to elicit reliable pitch percepts when 32 channels were used. Indeed, filter slopes as steep as 144 dB/oct still produced substantially poorer performance with a 64-channel vocoder than was observed in our non-overlap conditions.

Our conclusions that at least 32 channels without spectral overlap, or 64 channels with very steep filter slopes (>72 dB/oct) are required for complex spectral pitch extraction, are at odds with earlier studies that examined the number of channels and degree of spectral resolution needed for melody perception (Kong et al. 2004; Crew et al. 2012; Fielden et al. 2015). This apparent discrepancy can be ascribed to the fact that we were able to rule out cues based on temporal-envelope pitch and spectral edge pitch, which could have allowed the participants in the earlier studies to perform the tasks without extracting the pitch corresponding to the F0 from the low-numbered harmonics. Thus, it seems that when complex spectral pitch extraction is required with an accuracy of one musical semitone, a surprisingly high degree of spectral resolution is necessary.

Comparison of Spectral Resolution Required for Complex Spectral Pitch and Speech Perception

Studies investigating the number of vocoder channels needed for speech perception have found that the perception of simple sentences can be achieved with as few as four spectral channels (Shannon et al. 1995; Loizou et al. 1999; Faulkner et al. 2001) and have reported no systematic increase in performance with increasing number of channels beyond about eight (Friesen et al. 2001; Smith et al. 2002). A higher number of channels can improve performance when the speech is presented in background noise (Dorman et al. 1998), but even here, performance tends to plateau with between 12 and 20 spectral channels. In terms of spectral overlap, speech perception in noise becomes poorer with increasing amounts of spectral overlap; however, using slopes steeper than about 50 dB/oct does not further improve performance (Bingabr et al. 2008). In contrast, our results for melodic pitch perception suggest that up to 64 channels are required, and that performance continues to improve with filter slopes up to and beyond 144 dB/octave.

It should be noted that some elements of speech perception, including prosody and speaker identification, rely to some extent on pitch perception. We are not aware of any studies that have systematically examined the effects of number of channels or filter slope for pitch-related aspects of speech. Although the variations in pitch with speech are generally very large compared to the subtle pitch changes associated with melodies (Patel et al. 1998), it is still likely that robust pitch perception in a speech context (particularly in the presence of background noise or competing speech) requires the extraction of pitch from spectrally resolved harmonics. It therefore seems likely that the high number of channels and steep filter slopes required for the present task would also be needed to convey reliable pitch information in a speech context.

Effects of Asymmetric Spread on Melody Perception, Pitch Matching, and Pitch Comparisons

One important conclusion from our study is that accurate pitch perception can be elicited, so long as one of the slopes of each band-pass filter is very steep. The question of which slope is more important, and how it affects the overall pitch, appeared to be answered differently in the different experiments. The results from experiment 4, based on melody discrimination, suggested that similar performance was achieved whether the low or high filter slopes were steep. In contrast, the pitch-matching results from experiment 5 suggested that pitch accuracy was greater when the lower filter slope was steeper. One way to reconcile these results is to assume that the steeper lower slope (which coincides with the steeper slope of the excitation pattern after cochlear filtering; e.g., Zwicker 1970; Glasberg and Moore 1990) does result in more accurate pitch representations, but that the pitch generated by the steep high slope is still sufficiently accurate for good performance in the melody discrimination task.

In terms of the actual pitch generated by the vocoded stimuli, the results from the pitch-matching tasks strongly suggest that the pitch corresponds to the peak frequency of the vocoder filters, regardless of the slopes. Here again, an apparently different conclusion is reached when considering the results from the pitch-comparison experiment, where the stimuli with the steeper lower slope were consistently judged higher in pitch than the stimuli with the steeper upper slope. This finding is consistent with earlier studies, showing that the percept of tones with the same F0s can differ based on changes in brightness produced by shifts in the position of the spectral centroid of the tones (Samson et al. 1997; Caclin et al. 2005; Fastl and Zwicker 2007; Allen and Oxenham 2014). In other words, although the pitch remained the same, as demonstrated by the pitch-matching results, listeners responded in the pitch-comparison task based on perceived changes in brightness.

Mechanisms of Pitch Perception

The discussion of pitch in our current study has mainly focused on the spectral (or place) pitch that is extracted via tonotopically mapped spectral information. However, pitch may also be extracted by a purely temporal code or a combined spectro-temporal code. The temporal code suggests that periodicity is coded based on a population of neurons firing synchronously in the phase with the acoustic waveform (e.g., Cariani and Delgutte 1996; Plack et al. 2006), while the combined spectro-temporal code proposes that the period of an acoustic waveform could be determined by an array of coincidence detectors that correlates instantaneous phase information between different cochlear channels (e.g., Loeb et al. 1983; Shamma and Klein 2000). Physiological studies in non-human mammals have found auditory-nerve phase locking to be present robustly up to 1–2 kHz, and then strongly degraded above 2–3 kHz, depending somewhat on the species (Palmer and Russell 1986). Studies in humans have attempted to infer a limit of temporal coding from indirect psychophysical tests, and have reached limits that range from 1.5 kHz, based on the limits of binaural phase sensitivity (Grantham and Wightman 1978), to 4–5 kHz, based on frequency difference limens (Moore 1973), to 8 kHz, based on the point at which frequency difference limens no longer become systematically worse (Moore and Ernst 2012). Some evidence suggests that it is possible to extract complex pitch in normal-hearing listeners using only frequency components above 8 kHz (Oxenham et al. 2011), which is considered well above the limits of temporal phase locking. This outcome suggests that spectral coding, by itself, can be used for accurate pitch perception. Interestingly, animal vocoder studies of pitch have suggested that pitch processing in non-human mammals largely depends on temporal-envelope processing of unresolved harmonics, and that the sharper tuning associated with the human cochlea may result in stronger pitch extraction cues through spectral processing (Shofner and Campbell 2012; Shofner and Chaney 2013; Shofner 2014). Taken together, these outcomes support the interpretation of our current results in terms of spectral coding, although the role of temporal information cannot be ruled out.

Implications for Current and Future Implantable Auditory Prostheses

In the current study, we used noise-excited envelope vocoders. This technique has been used in many previous studies to simulate the CI users' performance on auditory tasks (Shannon et al. 1995; Dorman et al. 1998; Loizou et al. 1999; Rosen et al. 1999; Friesen et al. 2001; Qin and Oxenham 2005; Bingabr et al. 2008) and is widely believed to simulate aspects of CI processing, such as poor frequency selectivity and loss of temporal fine structure information. In many circumstances, the performance of NH listeners under vocoded conditions has been shown to be comparable to performance on the same tasks by CI users (e.g., Oxenham and Kreft 2014). Current CIs have somewhere between 12–24 active channels. The current spread produced on average seems to be equivalent to filter slopes of between 8 and 24 dB/oct, depending on the stimulation mode (monopolar stimulation, or more focused strategies, such as partial-tripolar stimulation), and on the assumptions made to relate electrical spread in the cochlea to acoustical filter slopes (Bingabr et al. 2008; Oxenham and Kreft 2014). Given that the minimum number channels required for accurate complex spectral pitch appears to be at least 32, with filter slopes exceeding 72 dB/octave, it seems unlikely that current CI devices will meet these constraints, even with significant further technical innovations. Therefore, alternative methods may be required to restore complex spectral pitch perception via implants. Given that the electrode-neuron interface is still a major limiting factor in CIs and noting the extremely high resolution needed for extraction of spectral pitch cues, neural prostheses such as intraneural electric stimulation (Middlebrooks and Snyder 2007, 2010) or neurotrophic alternatives (Pinyon et al. 2014; Wise et al. 2016) could be potential solutions. With current CIs, it seems that pitch must continue to be conveyed via the less salient and less accurate pitch elicited by periodic temporal-envelope fluctuations.

ACKNOWLEDGEMENTS

This work was supported by the NIH grant R01DC005216 (A.J.O).

COMPLIANCE WITH ETHICAL STANDARDS

The University of Minnesota Institutional Review Board gave approval for this experiment.

Conflict of Interest The authors declare that they have no conflict of interest.

REFERENCES

ABBERTON E, FOURCIN AJ (1978) Intonation and speaker identification. Lang Speech 21:305–318. doi:10.1177/002383097802100405

ALLEN EJ, OXENHAM AJ (2014) Symmetric interactions and interference between pitch and timbre. J Acoust Soc Am 135:1371–1379. doi:10.1121/1.4863269

AREHART KH (1994) Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners. J Acoust Soc Am 95:3574–3585. doi:10.1121/1.409975

BERNSTEIN JGW, OXENHAM AJ (2006) The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss. J Acoust Soc Am 120:3929–3945. doi:10.1121/1.2372452

BINGABR M, ESPINOZA-VARAS B, LOIZOU PC (2008) Simulating the effect of spread of excitation in cochlear implants. Hear Res 241:73–79. doi:10.1016/j.heares.2008.04.012

BURNS EM, VIEMEISTER NF (1976) Nonspectral pitch. J Acoust Soc Am 60:863–869. doi:10.1121/1.381166

CACLIN A, MCADAMS S, SMITH BK, WINSBERG S (2005) Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. J Acoust Soc Am 118:471–482. doi:10.1121/1.1929229

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

CARLYON RP (1996) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. J Acoust Soc Am 99:517–524. doi:10.1121/1.414510

CARLYON RP, DEEKS JM, MCKAY CM (2010) The upper limit of temporal pitch for cochlear-implant listeners: stimulus duration, conditioner pulses, and the number of electrodes stimulated. J Acoust Soc Am 127:1469–1478. doi:10.1121/1.3291981

COSENTINO S, CARLYON RP, DEEKS JM ET AL (2016) Rate discrimination, gap detection and ranking of temporal pitch in cochlear implant users. J Assoc Res Otolaryngol 17:371–382. doi:10.1007/s10162-016-0569-5

CREW JD, GALVIN JJ, FU Q-J (2012) Channel interaction limits melodic pitch perception in simulated cochlear implants. J Acoust Soc Am 132:EL429–EL435. doi:10.1121/1.4758770

DORMAN M, LOIZOU PC, FITZKE J, TU Z (1998) The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels. J Acoust Soc Am 104:3583–3585. doi:10.1121/1.423940

DUDLEY H (1939) Remaking Speech. J Acoust Soc Am 11:169–177. doi:10.1121/1.1916020

EWERT SD (2013) AFC—a modular framework for running psychoacoustic experiments and computational perception models. Proc Conf Acoust AIA-DAGA 2013:1326–1329

FASTL H, ZWICKER E (2007) Psychoacoustics—facts and models. Springer, Berlin

FAULKNER A, ROSEN S, WILKINSON L (2001) Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. Ear Hear 22:431–438

FIELDEN CA, KLUK K, BOYLE PJ, MCKAY CM (2015) The perception of complex pitch in cochlear implants: a comparison of monopolar and tripolar stimulation. J Acoust Soc Am 138:2524–2536. doi:10.1121/1.4931910

FRIESEN LM, SHANNON RV, BASKENT D, WANG X (2001) Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J Acoust Soc Am 110:1150–1163. doi:10.1121/1.1381538

FU Q-J, NOGAKI G (2005) Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J Assoc Res Otolaryngol 6:19–27. doi:10.1007/s10162-004-5024-3

GALVIN JJ, FU Q-J, NOGAKI G (2007) Melodic contour identification by cochlear implant listeners. Ear Hear 28:302–319. doi:10.1097/01.aud.0000261689.35445.20

GLASBERG BR, MOORE BCJ (1986) Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am 79:1020–1033. doi:10.1121/1.393374

GLASBERG BR, MOORE BCJ (1990) Derivation of auditory filter shapes from notched-noise data. Hear Res 47:103–138. doi:10.1016/0378-5955(90)90170-T

GRANTHAM DW, WIGHTMAN FL (1978) Detectability of varying interaural temporal differences. J Acoust Soc Am 63:511–523. doi:10.1121/1.381751

GREENWOOD DD (1990) A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am 87:2592–2605. doi:10.1121/1.399052

HIGHNAM C, MORRIS V (1987) Linguistic stress judgments of language learning disabled students. J Commun Disord 20:93–103. doi:10.1016/0021-9924(87)90001-3

HOUTSMA AJM, SMURZYNSKI J (1990) Pitch identification and discrimination for complex tones with many harmonics. J Acoust Soc Am 87:304–310. doi:10.1121/1.399297

KOHLRAUSCH A, HOUTSMA AJM (1992) Pitch related to spectral edges of broadband signals [and discussion]. Philos Trans R Soc Lond Ser B Biol Sci 336:375–382. doi:10.1098/rstb.1992.0071

KONG Y-Y, CRUZ R, JONES AJ, ZENG F-G (2004) Music perception with temporal cues in acoustic and electric hearing. Ear Hear 25(2):173–185

KONG Y-Y, DEEKS JM, AXON PR, CARLYON RP (2009) Limits of temporal pitch in cochlear implants. J Acoust Soc Am 125:1649–1657. doi:10.1121/1.3068457

LIEBERMAN P, MICHAELS SB (1962) Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. J Acoust Soc Am 34:922–927. doi:10.1121/1.1918222

LOEB GE, WHITE MW, MERZENICH MM (1983) Spatial cross-correlation. Biol Cybern 47:149–163. doi:10.1007/BF00337005

LOIZOU PC, DORMAN M, TU Z (1999) On the number of channels needed to understand speech. J Acoust Soc Am 106:2097–2103. doi:10.1121/1.427954

LUO H, WANG Y, POEPPEL D, SIMON JZ (2007) Concurrent encoding of frequency and amplitude modulation in human auditory cortex: encoding transition. J Neurophysiol 98:3473–3485. doi:10.1152/jn.00342.2007

MACHEREY O, DEEKS JM, CARLYON RP (2011) Extending the limits of place and temporal pitch perception in cochlear implant users. J Assoc Res Otolaryngol 12:233–251. doi:10.1007/s10162-010-0248-x

MACMILLAN NA, CREELMAN CD (2004) Detection theory: a user's guide. Psychology Press

MCDERMOTT HJ (2004) Music perception with cochlear implants: a review. Trends Amplif 8:49–82. doi:10.1177/108471380400800203

McDermott JH, Lehr AJ, Oxenham AJ (2008) Is relative pitch specific to pitch? Psychol Sci 19:1263–1271. doi:10.1111/j.1467-9280.2008.02235.x

Micheyl C, Keebler MV , Oxenham AJ (2010) Pitch perception for mixtures of spectrally overlapping harmonic complex tones. J Acoust Soc Am 128(1):257-269. doi:10.1121/1.3372751

Middlebrooks JC, Snyder RL (2007) Auditory prosthesis with a penetrating nerve array. J Assoc Res Otolaryngol 8:258–279. doi:10.1007/s10162-007-0070-2

Middlebrooks JC, Snyder RL (2010) Selective electrical stimulation of the auditory nerve activates a pathway specialized for high temporal acuity. J Neurosci 30:1937–1946. doi:10.1523/JNEUROSCI.4949-09.2010

Moore BCJ (1973) Frequency difference limens for short-duration tones. J Acoust Soc Am 54:610–619. doi:10.1121/1.1913640

Moore BCJ, Ernst SMA (2012) Frequency difference limens at high frequencies: evidence for a transition from a temporal to a place code. J Acoust Soc Am 132:1542–1547. doi:10.1121/1.4739444

Moore BCJ, Gockel HE (2011) Resolvability of components in complex tones and implications for theories of pitch perception. Hear Res 276:88–97. doi:10.1016/j.heares.2011.01.003

Moore BCJ, Huss M, Vickers DA et al (2000) A test for the diagnosis of dead regions in the cochlea. Br J Audiol 34:205–224. doi:10.3109/03005364000000131

Oxenham AJ (2012) Pitch perception. J Neurosci 32:13335–13338. doi:10.1523/JNEUROSCI.3815-12.2012

Oxenham AJ, Kreft HA (2014) Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. Trends Hear 18:2331216514553783. doi:10.1177/2331216514553783

Oxenham AJ, Micheyl C, Keebler MV et al (2011) Pitch perception beyond the traditional existence region of pitch. Proc Natl Acad Sci U S A 108:7629–7634. doi:10.1073/pnas.1015291108

Palmer AR, Russell IJ (1986) Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hear Res 24:1–15. doi:10.1016/0378-5955(86)90002-X

Patel AD, Peretz I, Tramo M, Labreque R (1998) Processing prosodic and musical patterns: a neuropsychological investigation. Brain Lang 61:123–144. doi:10.1006/brln.1997.1862

Pinyon JL, Tadros SF, Froud KE et al (2014) Close-field electroporation gene delivery using the cochlear implant electrode array enhances the bionic ear. Sci Transl Med 6:233ra54–233ra54. doi:10.1126/scitranslmed.3008177

Plack CJ, Oxenham AJ, Fay RR, Popper AN (2006) Pitch: neural coding and perception. Springer Science & Business Media

Plomp R (1967) Pitch of complex tones. J Acoust Soc Am 41:1526–1533. doi:10.1121/1.1910515

Pressnitzer D, Patterson RD, Krumbholz K (2001) The lower limit of melodic pitch. J Acoust Soc Am 109:2074–2084. doi:10.1121/1.1359797

Qin MK, Oxenham AJ (2005) Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. Ear Hear 26:451–460

Ritsma RJ (1967) Frequencies dominant in the perception of the pitch of complex sounds. J Acoust Soc Am 42:191–198. doi:10.1121/1.1910550

Rosen S, Faulkner A, Wilkinson L (1999) Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. J Acoust Soc Am 106:3629–3636. doi:10.1121/1.428215

Rosen S, Souza P, Ekelund C, Majeed AA (2013) Listening to speech in a background of other talkers: effects of talker number and noise vocoding. J Acoust Soc Am 133:2431–2443. doi:10.1121/1.4794379

Samson S, Zatorre RJ, Ramsay JO (1997) Multidimensional scaling of synthetic musical timbre: perception of spectral and temporal characteristics. Can J Exp Psychol Can Psychol Exp 51:307–315. doi:10.1037/1196-1961.51.4.307

Sayles M, Winter IM (2008) Reverberation challenges the temporal representation of the pitch of complex sounds. Neuron 58:789–801. doi:10.1016/j.neuron.2008.03.029

Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J Acoust Soc Am 95:3529–3540. doi:10.1121/1.409970

Shamma S, Klein D (2000) The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. J Acoust Soc Am 107:2631–2644. doi:10.1121/1.428649

Shannon RV, Zeng F-G, Kamath V et al (1995) Speech recognition with primarily temporal cues. Science 270:303

Shofner WP (2014) Perception of degraded speech sounds differs in chinchilla and human listeners. J Acoust Soc Am 135:2065–2077. doi:10.1121/1.4867362

Shofner WP, Campbell J (2012) Pitch strength of noise-vocoded harmonic tone complexes in normal-hearing listeners. J Acoust Soc Am 132:EL398–EL404. doi:10.1121/1.4757697

Shofner WP, Chaney M (2013) Processing pitch in a nonhuman mammal (*Chinchilla laniger*). J Comp Psychol 127:142–153. doi:10.1037/a0029734

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90. doi:10.1038/416087a

Stickney GS, Zeng F-G, Litovsky R, Assmann PF (2004) Cochlear implant speech recognition with speech maskers. J Acoust Soc Am 116:1081–1091. doi:10.1121/1.1772399

Titze IR (1989) Physiologic and acoustic differences between male and female voices. J Acoust Soc Am 85:1699–1707. doi:10.1121/1.397959

Wise AK, Tan J, Wang Y et al (2016) Improved auditory nerve survival with nanoengineered supraparticles for neurotrophin delivery into the deafened cochlea. PLoS One 11:e0164867. doi:10.1371/journal.pone.0164867

Zeng F-G, Tang Q, Lu T (2014) Abnormal pitch perception produced by cochlear implant stimulation. PLoS One 9:e88662. doi:10.1371/journal.pone.0088662

Zwicker E (1970) Masking and psychological excitation as consequences of the ear's frequency analysis. Freq Anal Period Detect Hear:376–394