Encoding pitch contours using current steering

Xin Luo[a)]

Department of Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive,
West Lafayette, Indiana 47907

David M. Landsberger, Monica Padilla, and Arthi G. Srinivasan

Division of Communication and Auditory Neuroscience, House Ear Institute, 2100 West Third Street,
Los Angeles, California 90057

This study investigated cochlear implant (CI) users' ability to perceive pitch cues from time-varying virtual channels (VCs) to identify pitch contours. Seven CI users were tested on apical, medial, and basal electrode pairs with stimulus durations from 100 to 1000 ms. In one stimulus set, 9 pitch contours were created by steering current between the component electrodes and the VC halfway between the electrodes. Another stimulus set only contained 3 pitch contours (flat, falling, and rising). VC discrimination was also tested on the same electrodes. The total current level of dual-electrode stimuli was linearly interpolated between those of single-electrode stimuli to minimize loudness changes. The results showed that pitch contour identification (PCI) scores were similar across electrode locations, and significantly improved at longer durations. For durations longer than 300 ms, 2 subjects had nearly perfect 9-contour identification, and 5 subjects perfectly identified the 3 basic contours. Both PCI and VC discrimination varied greatly across subjects. Cumulative d' values for VC discrimination were significantly correlated with 100-, 200-, and 500-ms PCI scores. These results verify the feasibility of encoding pitch contours using current steering, and suggest that identification of such pitch contours strongly relies on CI users' sensitivity to VCs. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3474237]

I. INTRODUCTION

Accurate pitch perception is important for human speech communication. In tonal languages such as Chinese, lexically meaningful tonal patterns are characterized by different pitch contours (e.g., Howie, 1976). In non-tonal languages such as English, pitch cues are used to signal stress, intonation, and speech segmentation (e.g., Lehiste, 1970). In addition to linguistic functions, pitch cues also convey talker indexical information (gender, identity, age, accent, etc.) and talker emotional state (happy, angry, sad, etc.). Moreover, human listeners use pitch cues to perform auditory scene analysis, or more specifically, segregate target speech from competing talker backgrounds (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; Culling and Darwin, 1994). Accurate pitch perception is also important for music appreciation, particularly for melody identification, which is related to discrimination between different pitch contours.

There are two underlying theories describing pitch perception in normal-hearing (NH) listeners: temporal and place theories (Licklider, 1951). According to the temporal theory of pitch coding, auditory nerve fibers fire at a rate that follows the input frequency (i.e., phase locking), at least for frequencies up to about 2 kHz (Johnson, 1980). Thus, the auditory system may use the time intervals between spikes to determine the input frequency. According to the place theory of pitch coding, high frequencies cause maximal basilar membrane displacement and excite auditory nerve fibers mostly near the basal end of the cochlea, while low frequencies cause maximal activity near the apical end of the cochlea. This tonotopic frequency-to-place mapping may also be used to determine the input frequency.

Cochlear implant (CI) processors exploit both place and temporal mechanisms to recover pitch perception, but with only limited success. Temporal pitch may be encoded by the rate and envelope of electric pulsatile stimulation (e.g., McKay and Carlyon, 1999). However, temporal pitch perception with CIs typically has an upper limit of about 300 Hz (e.g., Shannon, 1983; Zeng, 2002), although it has been shown with some subjects that temporal pitch can be perceived up to 1000 Hz (e.g., Townshend *et al.*, 1987; Wilson *et al.*, 2000; Landsberger and McKay, 2005). Place pitch coding is provided with CIs by stimulating different electrodes along the array. Typically, CI users perceive higher pitches when more basally located electrodes are stimulated (Nelson *et al.*, 1995), consistent with the tonotopic mapping in normal hearing. However, CI users' spectral resolution is greatly limited by the small number of implanted electrodes, the broad bandwidths of analysis filters, and the channel interaction created by broad patterns of electrical stimulation. With the limited pitch cues provided by CIs, it is not surprising that CI users perform poorly in pitch-related listening tasks (e.g., Luo *et al.*, 2007, 2008; Peng *et al.*, 2008; McDermott, 2004).

Current steering has been proposed to increase the number of distinctive pitch percepts beyond the limited number of electrodes available in commercially available CIs. This

---

[a)]Author to whom correspondence should be addressed. Electronic mail: luo5@purdue.edu

TABLE I. Subject demographic details.

Subject	Age	Gender	Etiology	Prosthesis	Strategy	Years with prosthesis
C1	76	M	Unknown	CII	HiRes-P w/Fidelity 120	7
C3	53	F	Genetic	HiRes90K	HiRes-S w/Fidelity 120	4
C4	62	F	Cochlear otosclerosis	CII	HiRes-S	3.8
C7	59	F	Fever+streptomycin	CII	HiRes-P w/Fidelity 120	3
C14	44	M	Maternal rubella	HiRes90k	HiRes-P w/Fidelity 120	4
C15	57	F	Genetic (otosclerosis labyrinthine)	HiRes90k	HiRes-S	4.8
C16R	56	F	Unknown	HiRes90k	HiRes-P w/Fidelity 120	1.7
C16L	56	F	Unknown	HiRes90K	HiRes-P w/Fidelity 120	0.7

technology adjusts the weighting of in-phase current delivered simultaneously to two adjacent electrodes (for an overview of current steering research, see Bonham and Litvak, 2008). As the current weighting is adjusted, the peak of the excitation pattern is steered to different locations between the two physical electrodes, eliciting distinctive intermediate pitch percepts and providing "virtual channels" (VCs) for information transmission. On average, CI users can perceive 5–7 VCs between each pair of adjacent electrodes (Donaldson et al., 2005; Firszt et al., 2007). Donaldson et al. (2005) also found that as long as the total current delivered to the two electrodes was kept constant or slightly increased (by <0.5 dB), the simultaneous dual-electrode stimuli were perceived as having similar loudness to the single-electrode stimuli. A speech processing strategy (Fidelity 120) has been developed for the Advanced Bionics cochlear implant system to provide spectral fine structure using current steering. Seven VCs are created between each pair of adjacent electrodes. During each stimulation cycle, the spectral peak within each of the 15 frequency bands is located to decide the place of the stimulated VC. Although CI users prefer Fidelity 120 over traditional strategies using only physical electrodes for music and everyday listening (Firszt et al., 2009), speech (including tone) recognition shows limited improvement with Fidelity 120 (Buechner et al., 2008; Firszt et al., 2009; Han et al., 2009). The apparent discrepancy between psychophysics and speech outcomes suggests that it may be necessary to rethink how to optimally utilize VCs in a multi-channel speech context.

In summary, studies have shown that most CI users can perceive intermediate pitch percepts elicited by current steering. A possible benefit of VC stimulation is better representation of dynamic cues in speech. The present study tests the feasibility of using VCs to encode pitch changes over time via a place pitch contour. Stimuli were continuously-steered currents on a pair of adjacent electrodes. Loudness was controlled within each pitch contour. Pitch contour identification (PCI) was measured to examine CI users' ability to perceive place pitch changes associated with time-varying VCs. Three electrode pairs were tested for each CI subject (representing the apical, medial, and basal portions of the electrode array) to see if there is any performance difference related to different electrode positions. In one stimulus set, a total of 9 pitch contours were included, which resemble the melodic contours developed by Galvin et al. (2007). In the other stimulus set, only 3 pitch contours (flat, falling, and rising)

were included, which resemble the pitch patterns of speech intonations and Chinese tones. Each stimulus set was tested with durations from 100 to 1000 ms, which cover the durations of various speech segments, from syllables to short phrases. PCI scores were compared to VC discrimination data to determine if the performance on the two tasks was related and if VC discrimination is the primary ability underlying PCI. It was hypothesized that better sensitivity to VCs would be predictive of higher-level PCI performance.

II. METHODS

A. Subjects

Seven postlingually deafened adult users of the Advanced Bionics Clarion CII or HiRes 90 K CI devices participated in this study. Subject C16 was bilaterally implanted with HiRes 90 K CIs. C16's performance with each implant was measured, so a total of 8 implanted ears were tested. Subjects used either the HiRes or the Fidelity 120 speech processing strategy in their clinically assigned speech processors. Table I shows the demographic details for the subjects. This study was reviewed and approved by the local IRB committees. All subjects provided informed consent and were compensated for their participation.

B. Pitch contour identification

1. Stimuli

Stimuli were delivered in monopolar mode to each subject's apical, medial, and basal electrode pairs using the Bionic Ear Data Collection System (Advanced Bionics, 2005). All subjects except for C1 were tested on electrode pairs 2–3, 7–8, and 13–14. Subject C1 was tested on electrode pairs 3–4, 8–9, and 14–15, because previous testing using partially tripolar stimulation suggested that for C1, electrode 7 had a poor electrode-neuron interface according to the arguments proposed by Bierer and Faulkner (2010). Subjects' loudness growth on each tested electrode was first estimated using the ascending method of limits. To measure loudness growth, a single physical electrode was stimulated with 300-ms, 1000-pulses/second, 42-μs/phase biphasic pulse trains. Subjects indicated the perceived loudness according to a 10-point scale from 1 ("barely audible") to 10 ("too loud"); this loudness scale is provided by Advanced Bionics to assist audiologists in CI programming. The stimulation level started from 5 μA and was gradually increased

Luo et al.: Virtual-channel pitch coding

in 5-μA steps until reaching the 8th ("maximally comfort-able") level. The current levels which corresponded to 1 ("barely audible"), 6 ("most comfortable"), and 8 ("maxi-mally comfortable") were recorded. For each electrode pair, the stimuli on the two component electrodes were loudness balanced with each other, using the method of adjustment. During the loudness-balance procedure, the most comfort-able stimulus on the apical electrode (as recorded during the loudness growth measurements) was played as a fixed refer-ence stimulus, and was followed by a comparison stimulus on the basal electrode using an inter-stimulus interval of 300 ms. The reference and comparison stimuli were alternated while subjects adjusted the level of the comparison stimulus in 1-μA steps using a knob (Griffin PowerMate). Subjects were asked to adjust the level of the second stimulus until it was the same loudness as the fixed-level stimulus. A mini-mum of three measurements per loudness balance were av-eraged to ensure reliable loudness-balanced levels.

Pitch contours were created between each tested elec-trode pair. A pitch contour consisted of a 1000-pulses/second VC pulse train. The proportion of current delivered to the basal electrode [α; consistent with the definition in previous studies (e.g., Donaldson et al., 2005; Bonham and Litvak, 2008)] was varied for each pulse during the stimulus. The first pulse for each stimulus was either only on the apical electrode ($\alpha=0$), the basal electrode ($\alpha=1$), or on a VC half-way between the two electrodes ($\alpha=0.5$). Similarly, the last pulse for each stimulus was presented at $\alpha=0$, 0.5, or 1. The pulse in the temporal center of the stimulus duration was always presented at $\alpha=0.5$. The value of α for each of the remaining pulses was linearly interpolated as a function of time. Nine different contours were created using the 3 differ-ent starting and 3 different ending values for α. For a de-tailed summary of the 9 contour stimuli, refer to Table II.

To minimize loudness variations within each pitch con-tour, the total current level (I) delivered to the two compo-nent electrodes of the electrode pair was changed as a linear function of α. Donaldson et al. (2005) measured the total current levels needed to produce equal loudness for simulta-neous, dual-electrode stimuli with different α values. They found that equal-loudness levels as a function of α (i.e., the equal-loudness contour) could be approximated using linear interpolation between the two end points ($\alpha=0$ and $\alpha=1$). The current levels required for simultaneous, dual-electrode stimuli were close to those for single-electrode stimuli, with differences less than 0.5 dB. Similar patterns of loudness-balance data were also found for the present subjects in a separate study (Landsberger and Srinivasan, 2009; data col-lected but not published). Based on these observations, the total current level (I) of the pitch contours in the present study was specified by time-varying α as follows:

$$I(t) = I_1 + (I_2 - I_1) \times \alpha(t), \tag{1}$$

where t represents time, $\alpha(t)$ was defined in Table II, I_1 is the most comfortable stimulation level (in μA) on the apical electrode, and I_2 is the loudness-balanced stimulation level (in μA) on the basal electrode (derived from the aforemen-tioned loudness-balance procedure). According to the short-term temporal-integration loudness model (McKay and Mc-

TABLE II. Pitch contour details.

Contour Name	Contour Shorthand	Starting α	Middle α	Ending α	Contour Drawing
Falling	Fa	1	0.5	0	
Falling-Flat	Fa-Fl	1	0.5	0.5	
Falling-Rising	Fa-Ri	1	0.5	1	
Flat-Falling	Fl-Fa	0.5	0.5	0	
Flat	Fl	0.5	0.5	0.5	
Flat-Rising	Fl-Ri	0.5	0.5	1	
Rising-Falling	Ri-Fa	0	0.5	0	
Rising-Flat	Ri-Fl	0	0.5	0.5	
Rising	Ri	0	0.5	1	

Dermott, 1998), such time-varying electric stimuli should generate roughly constant loudness throughout the pitch con-tours.

2. Procedures

PCI was measured for 2 stimulus sets (containing 3 and 9 contours) on 3 electrode pairs (apical, medial, and basal) with 5 stimulus durations (100, 200, 300, 500, and 1000 ms). Conditions within each stimulus set were tested in random order for each subject. In each run of the 9-contour tests, all 9 pitch contours (Fa, Fa-Fl, Fa-Ri, etc.) were randomly se-lected and presented to the subject 4 times, resulting in a total of 36 tokens per run. After listening to each stimulus, the subject was asked to identify the pitch contour by click-ing on one of the 9 response choices shown on a screen (i.e., a 9-alternative, forced-choice task). Each response choice was labeled with a line plot showing the pitch-variation pat-tern similar to those presented in Table II. In each run of the 3-contour tests, only the falling, rising, and flat pitch con-tours (stimuli 1, 5, and 9 in Table II) were randomly selected and presented to the subject 4 times, resulting in a total of 12 tokens per run. Similarly, the subject was asked to identify each pitch contour they heard by clicking on one of the 3 response choices shown on the screen (i.e., a falling line, a

J. Acoust. Soc. Am., Vol. 128, No. 3, September 2010

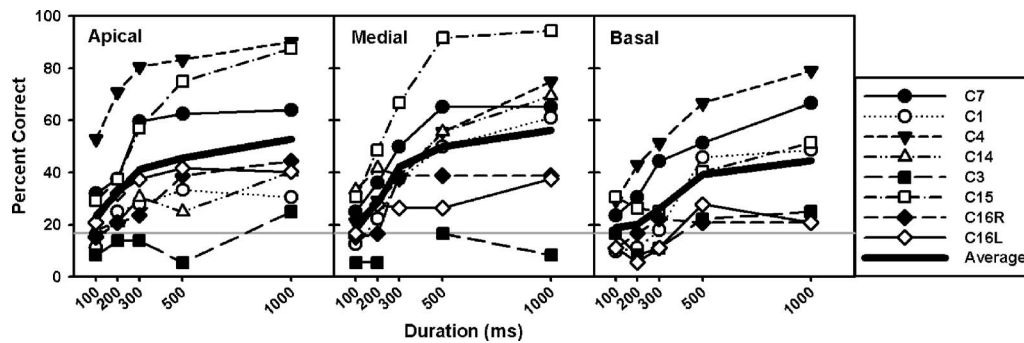Luo et al.: Virtual-channel pitch coding 1217

FIG. 1. 9-contour PCI scores as a function of stimulus duration. Different symbols and line types show the data for individual subjects. The thick solid lines represent the group mean. The gray horizontal lines indicate the 95% boundary for chance performance set by the binomial distribution (cumulative binomial probabilities). Performance on the apical, medial, and basal electrode pairs are shown in the left, middle, and right panels, respectively.

rising line, and a flat line). Subjects were instructed to ignore loudness variations and make their choices based on pitch changes. For both the 3- and 9-contour PCI tests, subject responses were recorded and percent correct scores were averaged across 2 runs. Before testing, subjects previewed each of the pitch contours as many times as they wanted. During the tests, there was no feedback.

C. Virtual channel discrimination

VC discrimination was also measured for each of the subjects. The electrode pairs and stimulation parameters were the same as used for the PCI tests, except that the phase duration of each pulse was 226 μs instead of 42 μs. VC discrimination data of subjects C1, C3, C4, and C7 were collected for a previous study (Landsberger and Srinivasan, 2009). The remaining subjects' (C14, C15, and C16) VC discrimination data were collected in the present study using the same procedure as in Landsberger and Srinivasan (2009). The VC discrimination procedure is briefly described here, and readers are encouraged to refer to Landsberger and Srinivasan (2009) for more details. Note that Landsberger and Srinivasan (2009) used the term monopolar virtual channel (MPVC) to describe VCs created in the manner described in this manuscript. For each electrode pair, VC stimuli with different α values (0.2, 0.4, 0.6, 0.8, and 1) were first loudness balanced to the most comfortable stimulus on the apical electrode ($\alpha = 0$). VC discrimination was then measured using a 3-interval, forced-choice (3IFC) task. On each trial, two intervals (randomly selected) contained the same α value, while the other interval contained a different α value. The subject was asked to identify which interval was different. To be consistent with the PCI tests, no feedback was provided for VC discrimination either. All pair-wise comparisons between the different α values were repeated 15 times. Finally, the percent correct discrimination scores between adjacent α values were converted to d′ scores, and the cumulative d′ value was calculated for each electrode pair. According to Hacker and Ratcliff (1979), 99% correct discrimination for the 3IFC task corresponds to a d′ value of 3.6, and the cumulative d′ value across the 6 α values can be as high as 18.

III. RESULTS

A. 9-contour PCI scores

Figure 1 shows the subjects' 9-contour PCI scores as a function of stimulus duration. There was large inter-subject variability in terms of the 9-contour PCI scores; one subject (C3) performed at chance level for all duration conditions, while others (e.g., C4 and C15) had nearly perfect performance on the apical and medial electrode pairs with stimulus durations longer than 500 ms. A two-way repeated measures analysis of variance (RM ANOVA) was performed to analyze the effects of stimulus duration and electrode location on the 9-contour PCI scores. Subject C14 was too tired and frustrated to complete the PCI tests on the basal electrode pair. Due to the missing data, C14 was not included in the two-way RM ANOVA. The analysis showed that the 9-contour PCI scores were not significantly different across the 3 electrode pairs [$F(2, 47) = 2.60$, $p = 0.12$], although the performance on the basal electrode pair tended to be poorer than that on the apical and medial electrode pairs. As the stimulus duration increased from 100 to 1000 ms, the mean 9-contour PCI scores significantly improved from 20% to 51% correct [$F(4, 47) = 27.06$, $p < 0.001$]. Post-hoc t-tests with the Bonferroni correction showed that except for the comparisons between adjacent duration conditions (e.g., 1000 vs. 500 ms, 500 vs. 300 ms, etc.), all pair-wise comparisons between duration conditions were significant ($p < 0.003$). There was no significant interaction between electrode location and stimulus duration [$F(8, 47) = 1.35$, $p = 0.24$].

A percent correct analysis of the 9-contour PCI data failed to reveal all of the information represented by the data. For example, flat-rising and rising-flat responses to a rising stimulus would both be counted as incorrect, even though they are better answers than a falling response. Thus, confusions among the 9 pitch contours were examined to better understand the effects of stimulus duration on PCI performance. Figure 2 shows the confusion matrix from all subjects across the 3 electrode pairs for the 100-ms stimuli. When examining the confusion matrix, it is of particular interest to separately look at the first part and the second part of the stimuli, effectively corresponding to two 3-contour stimuli. Table III shows the confusion matrices for the first

Response

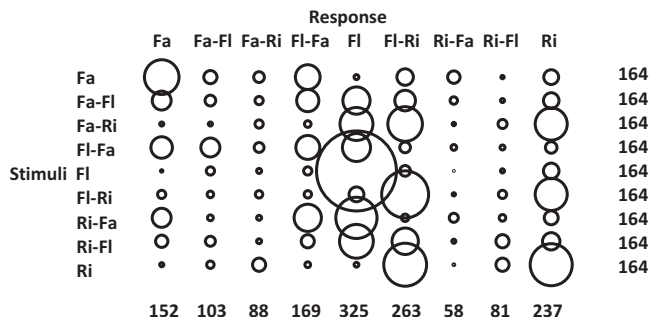| Stimuli	Fa	Fa-Fl	Fa-Ri	Fl-Fa	Fl	Fl-Ri	Ri-Fa	Ri-Fl	Ri	
Fa										164
Fa-Fl										164
Fa-Ri										164
Fl-Fa										164
Fl										164
Fl-Ri										164
Ri-Fa										164
Ri-Fl										164
Ri										164
	152	103	88	169	325	263	58	81	237	

FIG. 2. Confusion matrix from all subjects across the 3 electrode pairs for the 100-ms pitch contours. The percentages of pitch contour responses to target pitch contours are represented by the width of each bubble. The total number of each stimulus and response is shown on the right and bottom of the confusion matrix, respectively.

Response

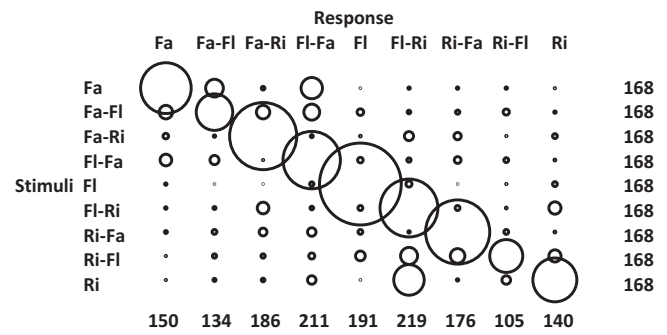| Stimuli	Fa	Fa-Fl	Fa-Ri	Fl-Fa	Fl	Fl-Ri	Ri-Fa	Ri-Fl	Ri	
Fa										168
Fa-Fl										168
Fa-Ri										168
Fl-Fa										168
Fl										168
Fl-Ri										168
Ri-Fa										168
Ri-Fl										168
Ri										168
	150	134	186	211	191	219	176	105	140	

FIG. 3. Confusion matrix from all subjects across the 3 electrode pairs for the 1000-ms pitch contours. The percentages of pitch contour responses to target pitch contours are represented by the width of each bubble. The total number of each stimulus and response is shown on the right and bottom of the confusion matrix, respectively.

and second parts of the 100-ms pitch contours. For the first part of the contours, subjects responded much more often with flat (51% of the time) than with rising and falling (25% and 23% of the time, respectively), and scored 38% correct. In contrast, for the second part of the contours, subjects' responses were less biased (flat, rising, and falling were selected for 34%, 39%, and 26% of the time, respectively), and the identification score (52% correct) was significantly higher than that for the first part of the contours [paired t-test: $t(7) = 2.58$, $p = 0.04$]. Overall, the Fl contour was best recognized (59% correct), followed by the Fl-Ri (35% correct) and Ri contours (31% correct), consistent with the aforementioned response biases. Although the overall 9-contour PCI scores significantly improved with increasing stimulus durations, the confusion patterns were largely similar across the 100-, 200-, and 300-ms duration conditions.

Figure 3 shows the confusion matrix from all subjects across the 3 electrode pairs for the 1000-ms stimuli, which is similar to that for the 500-ms stimuli. Table IV shows the confusion matrices for the first and second parts of the 1000-ms pitch contours. With the 1000-ms duration, subjects' responses were evenly distributed across flat, rising, and falling for both parts of the pitch contours (35%, 32%, and 33% of the time, respectively), and the first part of the contours was identified equally well (67% correct) as the second part of the contours (70% correct) [paired t-test: $t(7) = 0.83$, $p = 0.44$]. Compared to Fig. 2, there were fewer PCI confusions in Fig. 3. The remaining PCI confusions

were mostly among the Fa, Fa-Fl, and Fl-Fa contours or among the Ri, Ri-Fl, and Fl-Ri contours, which shared the same overall pitch changing direction (falling or rising) with each other. Of the 9 pitch contours, the Fl contour was best identified (80% correct), followed by the Fa-Ri (65% correct) and Ri-Fa contours (63% correct), both of which had opposite pitch changing directions (falling and rising) in the two parts of the contour.

B. 3-contour PCI scores

Figure 4 shows the subjects' 3-contour PCI scores as a function of stimulus duration. Due to the reduced number of choices and the increased chance performance level, the mean 3-contour PCI scores (e.g., 69% correct for 100-ms stimuli and 86% correct for 1000-ms stimuli) were much better than the 9-contour PCI scores. The 3-contour PCI scores were still largely variable across subjects, especially on the basal electrode pair and with durations shorter than 300 ms. The effects of stimulus duration and electrode location on the 3-contour PCI scores were analyzed using a two-way RM ANOVA. Similar to the 9-contour PCI scores, the 3-contour PCI scores were not significantly different across the 3 electrode pairs [$F(2, 56) = 3.14$, $p = 0.08$], but were significantly better with longer stimulus durations [$F(4, 56) = 10.51$, $p < 0.001$]; there was no significant interaction between electrode location and stimulus duration [$F(8, 56) = 1.09$, $p = 0.38$]. Post-hoc t-tests with the Bonferroni cor-

TABLE III. Confusion matrices (i.e., percentages of responses to stimuli) from all subjects across the 3 electrode pairs for the first and second parts of the 100-ms pitch contours.

Stimuli/response	Fa	Fl	Ri
Percentages of responses to the first part of the contours			
Fa	28	47	25
Fl	22	57	21
Ri	20	50	30
Percentages of responses to the second part of the contours			
Fa	44	31	25
Fl	22	47	31
Ri	11	25	64

TABLE IV. Confusion matrices (i.e., percentages of responses to stimuli) from all subjects across the 3 electrode pairs for the first and second parts of the 1000-ms pitch contours.

Stimuli/response	Fa	Fl	Ri
Percentages of responses to the first part of the contours			
Fa	68	21	11
Fl	14	73	13
Ri	11	29	60
Percentages of responses to the second part of the contours			
Fa	73	18	9
Fl	20	59	21
Ri	13	9	78

J. Acoust. Soc. Am., Vol. 128, No. 3, September 2010
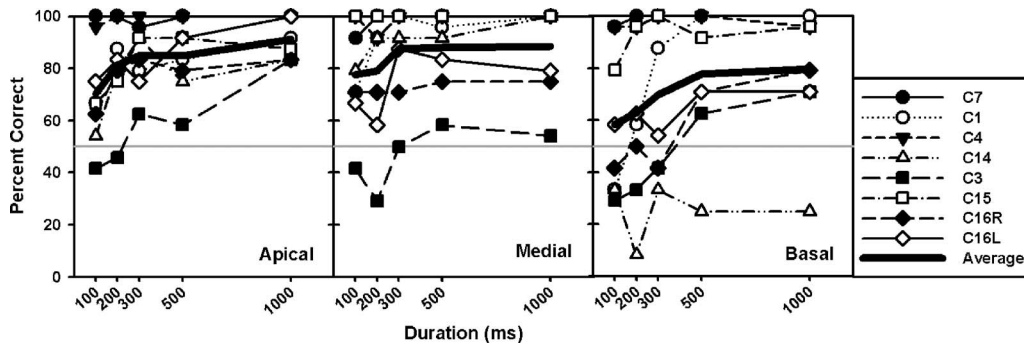
Luo et al.: Virtual-channel pitch coding 1219

FIG. 4. 3-contour PCI scores as a function of stimulus duration. Different symbols and line types show the data for individual subjects. The thick solid lines represent the group mean. The gray horizontal lines indicate the 95% boundary for chance performance set by the binomial distribution (cumulative binomial probabilities). Performance on the apical, medial, and basal electrode pairs are shown in the left, middle, and right panels, respectively.

rection showed that the 3-contour PCI scores were significantly better with the 300-, 500-, and 1000-ms durations than with the 100-ms duration ($p < 0.01$), and with the 1000-ms duration than with the 200-ms duration ($p < 0.01$).

Patterns of confusions among the 3 pitch contours (obtained from all subjects across the 3 electrode pairs) were similar with different stimulus durations, although the overall 3-contour PCI scores significantly improved with increasing stimulus durations. Subjects did not exhibit any strong response biases, but the flat contour was selected slightly less often (30% of the time) than the falling and rising contours (34% and 36% of the time, respectively). Interestingly, the flat pitch contour was also slightly less accurately identified (77% correct) than the falling and rising contours (78% and 84% correct, respectively).

C. Correlation between PCI scores and VC discrimination abilities

As was the case with the PCI scores, the ability to discriminate VCs also greatly varied from subject to subject (see Fig. 5). Individual subjects' cumulative d' values for VC discrimination (averaged across the 3 electrode pairs) ranged from 0.5 to 4.4. A one-way RM ANOVA showed that the cumulative d' values were not significantly different across the 3 electrode pairs $[F(2, 12) = 1.45, p = 0.27]$, consistent with the PCI scores. Correlation analyses were performed to determine if PCI scores could be predicted by a subject's

FIG. 5. Cumulative d' values for virtual channel discrimination on the apical, medial, and basal electrode pairs. Different symbols show the data for individual subjects. The horizontal lines represent the group mean.

ability to discriminate VCs. Figures 6 and 7 show the 9-contour and 3-contour PCI scores as a function of cumulative d' values, respectively. Note that, pitch contour identification was measured with different stimulus durations from 100 to 1000 ms and a 42-μs phase duration, while VC discrimination was measured with only the 300-ms stimulus duration and a 226-μs phase duration. Nevertheless, the cumulative d' values were significantly correlated with the 100-, 200-, and 500-ms 9-contour and 3-contour PCI scores (r ranged from 0.42 to 0.55, $p < 0.05$), and the correlations with the 300- and 1000-ms PCI scores approached, but did not reach, significance (r ranged from 0.29 to 0.43, p values ranged from 0.06 to 0.20). The 3-contour PCI scores were subject to ceiling effects (especially with durations longer than 300 ms). Accordingly, the cumulative d' values were more strongly correlated with the 9-contour PCI scores than with the 3-contour PCI scores (see Figs. 6 and 7 for correlation coefficients).

IV. DISCUSSION

In the present study, current steering was used to encode various pitch contours for CI users on apical, medial, and basal electrode pairs. With a stimulus duration that is typical for a word or a short phrase (i.e., 500 and 1000 ms), most CI users were able to perceive pitch changes associated with time-varying VCs and identify the target pitch contours with above-chance performance. For the 9-contour PCI tests, subjects C4 and C15 even scored ~90% correct on apical or medial electrode pair. For the 3-contour PCI tests, 5 out of the 8 subjects scored 100% correct on at least one of the electrode pairs. These results verify the feasibility of using current steering to encode pitch contours (in speech or music) for CI users. This pitch coding strategy does not intend to recover the exact pitch perception in acoustic hearing, but rather transmits the relative pitch variation patterns, which are important for tone, intonation, and melody recognition. Using this strategy, fundamental frequency changes in syllables or sentences can be encoded on apical electrodes (corresponding to lower frequencies). Formant frequency transitions may also be encoded on medial or basal electrodes (corresponding to higher frequencies). However, the durations of formant transitions in syllables are typically shorter than 100 ms. The finding that subjects performed more
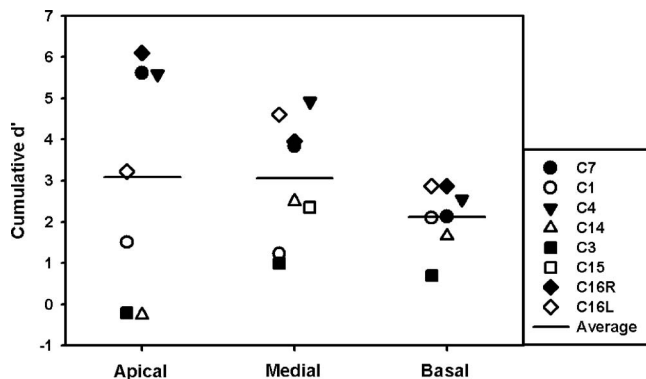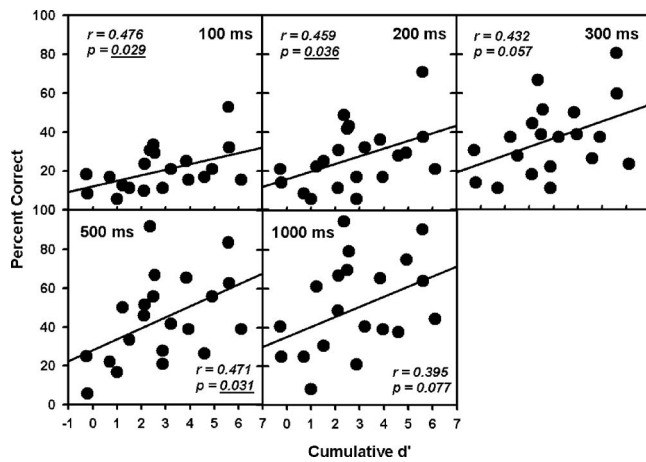
FIG. 6. 9-contour PCI scores as a function of cumulative d′ values for virtual channel discrimination. Different panels show the 9-contour PCI scores with different stimulus durations. The filled circles represent individual subjects' performance on individual electrode pairs. The solid lines show the linear regressions between the cumulative d′ values and the 9-contour PCI scores. Significant correlations are indicated by underlining the p values within each panel.
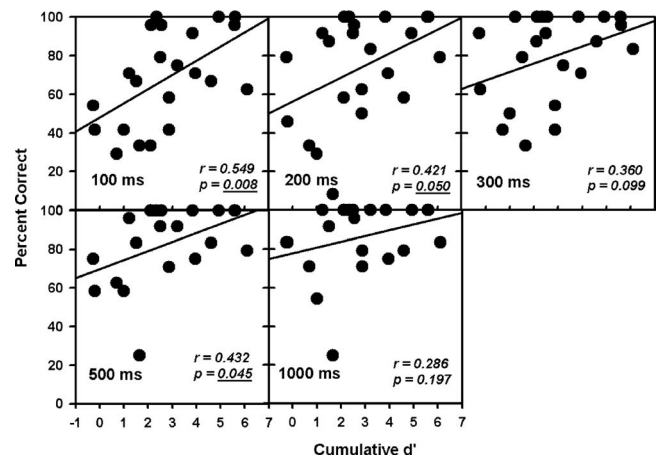

FIG. 7. 3-contour PCI scores as a function of cumulative d′ values for virtual channel discrimination. Different panels show the 3-contour PCI scores with different stimulus durations. The filled circles represent individual subjects' performance on individual electrode pairs. The solid lines show the linear regressions between the cumulative d′ values and the 3-contour PCI scores. Significant correlations are indicated by underlining the p values within each panel.

poorly on the PCI test with shorter durations suggests that they may have great difficulty identifying formant transitions. Using the same frequency map as in the Fidelity 120 strategy, electrode pair 2–3 would cover frequencies from 455 to 540 Hz, electrode pair 7–8 would cover 1076 to 1278 Hz, and electrode pair 13–14 would cover 3022 to 3590 Hz. In speech, fundamental and formant frequencies (F0, F1, and F2, etc.) have variation ranges including but not limited to those tested here, thus may be mapped to different electrodes with different amounts of current steering. For example, the F0 of a rising tone produced by a female talker can increase from 180 to 320 Hz, while the F2 of a transition from /d/ to /a/ produced by a male talker can decrease from 1630 to 1110 Hz. Future studies need to investigate the effects of frequency or tonotopic sweep extent on PCI results.

The present study measured CI users' identification of continuous pitch contours created by steering current on individual electrode pairs. It is worthwhile to compare our results with those of Galvin et al. (2007), who measured CI users' melodic contour identification (MCI) using stimuli presented through subjects' clinically assigned speech processors. In the study of Galvin et al. (2007), 5 discrete music notes (250 ms each) were used to create the same 9 pitch contours as in the present study; the frequency of the lowest note in each contour was varied to cover different frequency regions, and the interval between successive notes in each contour ranged from 1 to 5 semitones. In the present study, subjects were only provided place cues between two adjacent electrodes to identify the pitch contours, while the subjects in Galvin et al. (2007) may have attended to multi-electrode stimulation patterns (including both place cues across channels and temporal cues within channels) to identify the melodic contours. Despite these differences, the 1000-ms 9-contour PCI scores of the present subjects were in the same range (from 21% to 94% correct) as the MCI scores reported by Galvin et al. (2007); both studies showed a large inter-subject variability. Detailed performance patterns were also

similar between the two studies. For example, of the 9 pitch contours, the Fl contour was best identified and the Fa-Ri and Ri-Fa contours were identified with relatively high accuracy.

As a group, the present subjects showed similar PCI scores and VC discrimination data on the apical, medial, and basal electrode pairs. However, similar to the observations of Donaldson et al. (2005), performance varied across electrodes in some subjects. For example, C14 had perfect 3-contour PCI scores on the medial electrode pair, but only performed at chance level on the basal electrode pair. C14 did have lower cumulative d′ value for VC discrimination on the basal electrode pair (1.6) than on the medial electrode pair (2.5). However, the poorer VC sensitivity on the basal electrode pair cannot fully explain the corresponding chance-level PCI scores because electrodes 13 and 14 were nevertheless perceptually discriminable to C14. It is possible that although C14 could hear the difference between steady-state stimuli on electrodes 13 and 14, the subject was unsure about the relative pitch change in dynamic stimuli. This example suggests that even if current steering between adjacent electrodes is implemented on the whole electrode array, individual CI users may not have equal access to the encoded spectral details in different frequency ranges (e.g., pitch changes or formant transitions). Since electrode pairs with chance-level PCI scores are unlikely to provide additional spectral details, steering current only on electrode pairs with better-than-chance PCI scores may result in similar (if not better) performance as steering current on all electrode pairs.

Subjects' PCI scores significantly improved with increasing stimulus durations. Longer durations provided more detailed sampling of the same pitch variation range and allowed more time for intermediate place pitch perception, which may have helped CI users' discrimination of VC pitch cues, and improved the identification of pitch variation patterns. Similar to the present PCI scores, Luo et al. (2010) also found that CI users' temporal pitch perception (50- and

100-Hz AM frequency discrimination) significantly improved as the stimulus duration increased up to 5–10 modulation periods. It is unknown if PCI performance would be better with a higher stimulation rate, which provides more detailed pitch sampling within a fixed time duration.

For the 9-contour PCI tests with durations shorter than 300 ms, subjects were unable to reliably identify the first part of the contours (performance close to chance), and their responses were strongly biased toward flat. In contrast, the second part of the contours was identified with significantly higher accuracy (although still limited). With such short stimulus durations, both parts of the contours may have had ambiguous neural representations. It is plausible that the later-formed second-part neural representations may have been better remembered in short-term memory than the earlier-formed first-part neural representations, thus were better identified by CI users. It appears that the identification of the two parts of the contours was independent of each other, in light of their unrelated, different response patterns. When the stimulus duration was longer than 300 ms, both parts of the contours were equally well identified, suggesting that their neural representations were enhanced to comparable levels of robustness and were thus equally well remembered in short-term memory regardless of their time order.

In the 3-contour PCI tests, the identification scores were not significantly different between the rising and falling pitch contours. Interestingly, Gordon and Poeppel (2002) reported that NH listeners were significantly better at identifying upward frequency-modulated (FM) tone glides than downward FM tone glides when both glides spanned the same half-octave frequency range and the glide duration was shorter than 160 ms. In normal hearing, upward FM glides may compensate for the traveling wave delay between high and low input frequencies, thus may produce more synchronized neural firing than downward FM glides (Shore and Nuttall, 1985). The greater neural responses to upward FM glides may have resulted in better identification scores than those of downward FM glides. In electric hearing, the 0.85-mm electrode spacing on the Advanced Bionics' HiFocus electrode array covers a frequency range much smaller than half an octave [according to Greenwood's (1990) formula]. Moreover, there is no traveling wave delay between high and low frequencies (due to direct electric stimulation), thus neural responses to the falling and rising pitch contours may not be as different as in normal hearing, which may explain the similar identification scores for the two pitch contours.

Subjects' ability to identify pitch contours was significantly correlated with their sensitivity to VCs (in terms of cumulative d' values). Subjects with higher VC sensitivity may have perceived more distinctive pitches as current was steered between adjacent electrodes, which may have contributed to better PCI scores. However, there were also exceptions. For example, the subject with the highest cumulative d' value did not perform the best among all subjects. As mentioned before, the VC discrimination tests only required subjects to hear the difference between steady-state VCs, while in the PCI tests, subjects were required to identify the pitch patterns of time-varying VCs. Clearly, the latter had higher perceptual task demands, and its performance may not

be predicted by VC discrimination ability alone. Presumably, adequate resolution in place pitch as well as perceptual integration of pitch information across excitation place and time is needed for the PCI task. Besides, the present study only measured acute PCI performance. Targeted PCI training with feedback may be necessary for the subject with the highest cumulative d' value to better understand the difference between various pitch contours, and in turn, achieve better PCI scores. On the other hand, the significant correlation between PCI and VC discrimination suggests that it may be beneficial to use more focused tripolar or quadrupolar VCs instead of less-focused monopolar VCs to encode pitch contours. These current focusing techniques have been shown to improve VC discrimination ability (e.g., Landsberger and Srinivasan, 2009), thus are expected to provide more distinctive place cues for pitch contour identification.

In this study, only recipients of an Advanced Bionics implant system were tested. It is worth noting that the results from current steering are dependent on the geometry specific to the electrode array. The Nucleus straight and contour arrays have electrode spacing of approximately 0.7 mm (which is relatively similar to the 0.85-mm spacing on the Advanced Bionics' HiFocus electrode array). Nucleus devices only have one current source so they cannot provide a simultaneous VC in the same way implemented in this study. However, it has been shown with Nucleus devices that sequential stimulation of adjacent electrodes can create intermediate place pitches (McDermott and McKay, 1994; Kwon and van den Honert, 2006). The Nucleus CI24RE array has the ability to short two adjacent electrodes and effectively stimulate them simultaneously with an α of 0.5, providing a pitch intermediate to that of either component electrode (Busby *et al.*, 2008). The MED-EL PULSARCI[100] array has multiple current sources and therefore can be used to steer current between adjacent electrodes and provide pitches intermediate to those of component electrodes (Nobbe *et al.*, 2007). However, electrodes on the PULSARCI[100] array are 2.4 mm apart. Therefore the current overlap between adjacent electrodes should be much smaller than using adjacent electrodes in Advanced Bionics devices. It is therefore possible to generate place pitch contours for PULSARCI[100] users, but it is unknown how their ability to label place pitch glides would compare with the data presented in this manuscript.

It is worth noting that the present PCI results were obtained on single electrode pairs. In CI speech processing strategies such as Fidelity 120, speech cues are presented to multiple electrodes, with both temporal AM on individual electrodes and current steering between adjacent electrodes. To further test the feasibility of using current steering to encode pitch contours in real CI speech processors, future studies need to investigate CI users' ability to integrate place pitch cues (elicited from VCs) with temporal pitch cues (elicited from AM) within each channel, and their ability to integrate place pitch cues across multiple channels. Such integration may impair or enhance PCI performance depending on whether the integrated pitch cues are conflicting or cooperating with each other.

Advanced Bionics (**2005**). Bionic Ear Data Collection System, Version 1.17 Users Manual.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Bierer, J. A., and Faulkner, K. F. (**2010**). "Identifying cochlear implant channels with poor electrode-neuron interface: Partial tripolar, single-channel thresholds and psychophysical tuning curves," Ear Hear. **31**, 247–258.

Bonham, B. H., and Litvak, L. M. (**2008**). "Current focusing and steering: Modeling, physiology, and psychophysics," Hear. Res. **242**, 141–153.

Brokx, J. P., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Buechner, A., Brendel, M., Krüeger, B., Frohne-Büchner, C., Nogueira, W., Edler, B., and Lenarz, T. (**2008**). "Current steering and results from novel speech coding strategies," Otol. Neurotol. **29**, 203–207.

Busby, P. A., Battmer, R. D., and Pesch, J. (**2008**). "Electrophysiological spread of excitation and pitch perception for dual and single electrodes using the nucleus freedom cochlear implant," Ear Hear. **29**, 853–864.

Culling, J. F., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.

Donaldson, G. S., Kreft, H. A., and Litvak, L. (**2005**). "Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users," J. Acoust. Soc. Am. **118**, 623–626.

Firszt, J. B., Holden, L. K., Reeder, R. M., and Skinner, M. W. (**2009**). "Speech recognition in cochlear implant recipients: Comparison of standard HiRes and HiRes 120 sound processing," Otol. Neurotol. **30**, 146–152.

Firszt, J. B., Koch, D. B., Downing, M., and Litvak, L. (**2007**). "Current steering creates additional pitch percepts in adult cochlear implant recipients," Otol. Neurotol. **28**, 629–636.

Galvin, J. J., Fu, Q.-J., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear Hear. **28**, 302–319.

Gordon, M., and Poeppel, D. (**2002**). "Inequality in identification of direction of frequency change (up versus down) for rapid frequency modulated sweeps," ARLO **3**, 29–34.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d′ for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.

Han, D., Liu, B., Zhou, N., Chen, X., Kong, Y., Liu, H., Zheng, Y., and Xu, L. (**2009**). "Lexical tone perception with HiResolution and HiResolution 120 sound-processing strategies in pediatric Mandarin-speaking cochlear implant users," Ear Hear. **30**, 169–177.

Howie, J. M. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge), pp. 1–279.

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

Kwon, B. J., and van den Honert, C. (**2006**). "Dual-electrode pitch discrimination with sequential interleaved stimulation by cochlear implant users," J. Acoust. Soc. Am. **120**, EL1–EL6.

Landsberger, D. M., and McKay, C. M. (**2005**). "Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implants," J. Acoust. Soc. Am. **117**, 319–327.

Landsberger, D. M., and Srinivasan, A. G. (**2009**). "Virtual channel discrimination is improved by current focusing in cochlear implant recipients," Hear. Res. **254**, 34–41.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA), pp. 1–194.

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experimentia **7**, 128–134.

Luo, X., Fu, Q.-J., and Galvin, J. J. (**2007**). "Vocal emotion recognition by normal-hearing listeners and cochlear implant users," Trends Amplif. **11**, 301–315.

Luo, X., Fu, Q.-J., Wei, C.-G., and Cao, K.-L. (**2008**). "Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users," Ear Hear. **29**, 957–970.

Luo, X., Galvin, J. J., and Fu, Q.-J. (**2010**). "Effects of stimulus duration on amplitude modulation processing with cochlear implants," J. Acoust. Soc. Am. **127**, EL23–EL29.

McDermott, H. J. (**2004**). "Music perception with cochlear implants: A review," Trends Amplif. **8**, 49–82.

McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.

McKay, C. M., and Carlyon, R. P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am. **105**, 347–357.

McKay, C. M., and McDermott, H. J. (**1998**). "Loudness perception with pulsatile electrical stimulation: The effect of interpulse intervals," J. Acoust. Soc. Am. **104**, 1061–1074.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.

Nobbe, A., Schleich, P., Zierhofer, C., and Nopp, P. (**2007**). "Frequency discrimination with sequential or simultaneous stimulation in MED-EL cochlear implants," Acta Oto-Laryngol. **127**, 1266–1272.

Peng, S.-C., Tomblin, J. B., and Turner, C. W. (**2008**). "Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing," Ear Hear. **29**, 336–351.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.

Shore, S. E., and Nuttall, A. L. (**1985**). "High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts," J. Acoust. Soc. Am. **78**, 1286–1295.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

Wilson, B., Wolford, R., and Lawson, D. (**2000**). "The Sixth Quarterly Progress Report, NIH/NIDCD N01-DC-8–2105: Speech processors for auditory prostheses," http://www.rti.org/reports/capr/N01-DC-8-2105QPR06.pdf (Last viewed 6/30/2010).

Zeng, F.-G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

J. Acoust. Soc. Am., Vol. 128, No. 3, September 2010

Luo *et al.*: Virtual-channel pitch coding   1223