

Channel interaction limits melodic pitch perception in simulated cochlear implants

Joseph D. Crew^{a),b)}

*Department of Biomedical Engineering, University of Southern California,
Los Angeles, California 90089
jcrew@hei.org*

John J. Galvin III

*Department of Communication and Auditory Neuroscience, House Research Institute,
Los Angeles, California 90057
jgalvin@hei.org*

Qian-Jie Fu^{b)}

*Department of Biomedical Engineering, University of Southern California,
Los Angeles, California 90089
qfu@hei.org*

Abstract: In cochlear implants (CIs), melodic pitch perception is limited by the spectral resolution, which in turn is limited by the number of spectral channels as well as interactions between adjacent channels. This study investigated the effect of channel interaction on melodic contour identification (MCI) in normal-hearing subjects listening to novel 16-channel sinewave vocoders that simulated channel interaction in CI signal processing. MCI performance worsened as the degree of channel interaction increased. Although greater numbers of spectral channels may be beneficial to melodic pitch perception, the present data suggest that it is also important to improve independence among spectral channels.

© 2012 Acoustical Society of America

PACS numbers: 43.66.Ts, 43.66.Hg, 43.75.Cd [DD]

Date Received: August 17, 2012 **Date Accepted:** September 28, 2012

1. Introduction

Multi-channel cochlear implants (CIs) have restored the sensation of hearing to many profoundly deaf individuals. Most CI users are able to understand speech quite well in quiet listening conditions. However, CI performance in difficult listening tasks (e.g., speech understanding in noise, music perception) remains much poorer than normal hearing (NH) listeners, primarily due to the limited spectral resolution of the device (Gfeller *et al.*, 2002; McDermott, 2004). In CIs, spectral resolution is limited by the number of physically implanted electrodes (or virtual channels in the case of current steering), as well as by the amount of channel interactions between adjacent electrodes. In CIs, channel interaction can be due to unintended electric field interactions from nearby electrodes leading to an overlapping spread of excitation of the auditory nerve.

The effect of spectral resolution on speech understanding has been extensively studied in CI users (Friesen *et al.*, 2001; Fu and Nogaki, 2005; Luo *et al.*, 2007; Bingabr *et al.*, 2008). Friesen *et al.* (2001) found that, for NH subjects, speech performance in quiet and noise steadily improved as the number of channels was increased up to 20. For CI subjects, performance steadily improved up to seven to ten spectral

^{a)}Author to whom correspondence should be addressed.

^{b)}Also at: Department of Communication and Auditory Neuroscience, House Research Institute, Los Angeles, CA 90057.

channels, beyond which there was no improvement, presumably due to channel interaction. Luo *et al.* (2007) demonstrated that vocal emotion recognition, which relies strongly on voice pitch cues, worsened as the number of spectral channels was reduced, for both real CI users and NH subjects listening to CI simulations. Fu and Nogaki (2005) investigated the influence of channel interaction on speech perception in gated noise for NH subjects listening to acoustic CI simulations. They found that, for a wide range of gating frequencies, speech understanding in noise worsened as the amount of channel interaction increased. They also found that mean CI performance was most similar to that of NH subjects listening to four spectrally smeared channels, with the top CI performance similar to that of NH subjects listening to 8–16 spectrally smeared channels. Bingabr *et al.* (2008) simulated the spread of excitation in an acoustic CI simulation using a model based on neurophysiologic data. Again, effective spectral resolution was reduced as the amount of channel interaction was increased, resulting in poorer speech performance in quiet and in noise.

Limited spectral resolution has also been shown to negatively impact music perception. Kong *et al.* (2004) observed that decreasing the number of spectral channels influenced melody recognition. Less is known about the effects of channel interaction on music perception, specifically melodic pitch perception. CI users must extract pitch from coarse spectro-temporal representations and do not have access to fine structure cues that are important for melodic pitch and timbre perception. Indeed, pitch and timbre may sometimes be confounded in CI users. Galvin *et al.* (2008) found that, different from NH listeners, CI users' melodic contour identification (MCI) was significantly affected by instrument timbre. MCI performance was generally better with simpler stimulation patterns (e.g., organ) than with complex patterns (e.g., piano). It is unclear how channel interaction (which relates in some ways to broad stimulation patterns) may affect melodic pitch perception.

Most recent CI research and development has been directed at improving the number of spectral channels (e.g., virtual channels) rather than reducing channel interaction. The present study investigated the effect of channel interaction on melodic pitch perception, using the MCI task. NH subjects were tested while listening to novel 16-channel sinewave vocoders that simulated varying amounts of channel interaction. We hypothesized that, similar to previous CI speech studies, increased channel interaction would reduce melodic pitch salience, resulting in poorer MCI performance.

2. Methods

Twenty NH subjects (aged 23–48 years) participated in the experiment. For all subjects, pure-tone thresholds were less <20 dB for audiometric frequencies up to 8 kHz. Music pitch perception was measured using a MCI task (Galvin *et al.*, 2007). The nine contours were “rising,” “falling,” “flat,” “rising-flat,” “falling-flat,” “rising-falling,” “falling-rising,” “flat-rising,” and “flat-falling.” Each contour consisted of five musical notes 300 ms in duration with 300 ms between notes. The lowest note of each contour was A3 (220 Hz). The frequency spacing between successive notes in each contour was varied between one and three semitones. The source stimuli for the vocoding was a piano sound, created by musical instrument digital interface (MIDI) sampling and re-synthesis, as in Galvin *et al.* (2008) and Zhu *et al.* (2011).

The CI simulations were vocoded using sinewave carriers, rather than noise-bands. Although phoneme and sentence recognition performance has been shown to be similar with sinewave and noise-band CI simulations (Dorman *et al.*, 1997), sinewave simulations have been shown to better emulate real CI performance for pitch-related speech tasks (Luo *et al.*, 2007). Sinewave carriers offer better stimulation site specificity and better temporal envelope representation than noise-band carriers. One concern with sinewave-vocoding is the potential for additional pitch cues provided by sidebands resulting from amplitude modulation. However, in this study, if such sideband pitch cues were available, they were equally available across all channel interaction conditions (the parameter of concern).

The present study made use of a novel implementation of an acoustic CI simulation, in that it also simulated different amounts of channel interaction. Figure 1 shows a schematic representation of the simulated CI signal processing, which was implemented as follows. The source stimulus was fed into a bank of 16 bandpass filters equally spaced according to the frequency-to-place mapping of Greenwood (1990). The overall input frequency range was 188–7938 Hz, which corresponded to the default input frequency range of Cochlear Corp.’s (Sydney, NSW, Australia) Nucleus-24 device. The slowly varying envelope energy in each band was extracted via half-wave rectification followed by low-pass filtering (cutoff frequency of 160 Hz). Different degrees of channel interaction were simulated by adding variable amounts of temporal envelope information extracted across analysis bands to the envelope of a particular band. This was analogous to modifying the output filter slopes in noise-band vocoding. The amount of envelope information added to adjacent bands depended not only on the targeted degree of channel interaction, but also on the frequency distance between adjacent bands. The output filter slopes were 24, 12, or 6 dB/octave, simulating “slight,” “moderate,” and “severe” channel interaction, respectively. The temporal envelope from each band (including the targeted degree of channel interaction) was used to modulate a sinewave carrier whose frequency corresponded to the center frequency of the analysis band. The outputs were then summed and the resulting signal was normalized to have the same long-term root-mean-square amplitude of the input signal. Audio examples of the unprocessed stimuli and the vocoded stimuli with different amounts of channel interaction are given in Mm. 1.

Mm. 1. Audio example of the rising contour with two-semitone spacing: (1) Unprocessed, followed by the 16-channel CI simulation with (2) no, (3) slight, (4) moderate, and (5) severe channel interaction. This is a file of type “wav” (2247 kB).

Figure 2 shows frequency analysis for A3 (the lowest note in any of the contours), B3 (two semitones higher than A3), and C#4 (two semitones higher than B3), for unprocessed stimuli and stimuli processed by the CI simulation with no, slight or severe channel interaction. These three notes are the first three notes heard in audio demo Mm. 1. As the amount of channel interaction was increased, the difference in

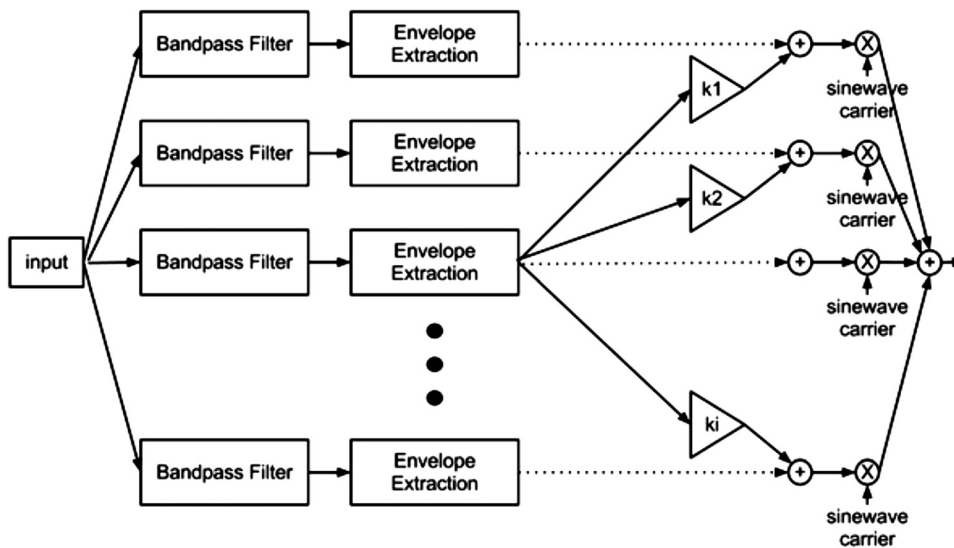


Fig. 1. Block diagram of the signal processing for the acoustic CI simulation. Temporal envelope information extracted from an analysis band (only one band is shown) is added to other bands with a gain of k_i , which corresponds to filter slope in dB/octave. The sinewave carriers were the center frequency of the frequency analysis band.

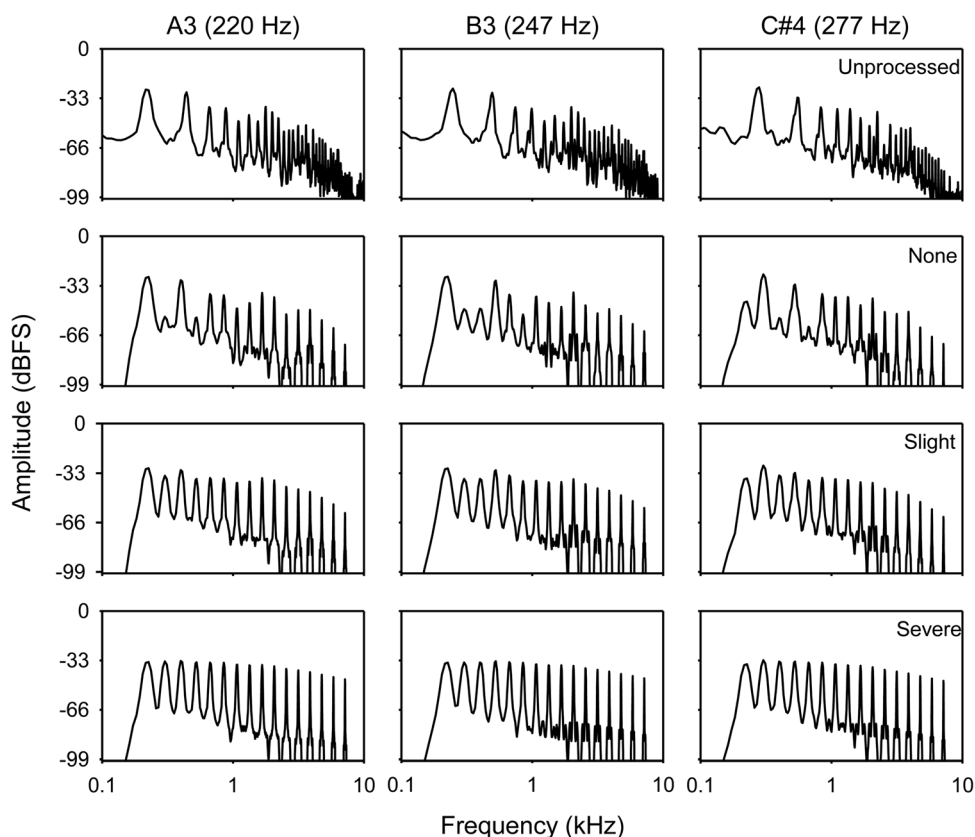


Fig. 2. Frequency analysis of experimental stimuli. From the left to right, the columns indicate different notes. The top row shows the unprocessed signal. Rows 2, 3, and 4 show stimuli processed by the CI simulations with no, slight, and severe channel interaction, respectively.

spectral envelope contrast or depth across notes was reduced. When the contrast is sufficiently reduced (bottom row), there appears to be little change in the spectral envelope across the three notes. With no channel interaction (row 2), the depth is sufficient to see the peaks of the spectral envelope shift across notes. However, the representation with the acoustic CI simulation, even with no channel interaction [row 2 in Fig. 2 and the second example in Mm. 1], is much poorer than for the unprocessed stimuli [top row of Fig. 2 and the first example in Mm. 1]. Figure 2 suggests that pitch cues across notes will be better preserved as the degree of channel interaction is reduced.

Each subject was initially tested using the unprocessed stimuli to familiarize the subject with the experiment and to verify that they were able to score above 90% correct for the MCI task. All subjects were tested while sitting in a sound-treated booth (IAC, Bronx, NY) and directly facing a single loudspeaker [Tannoy (Coatbridge, Scotland, UK) Reveal]. All stimuli were presented acoustically at 65 dBA. The four-channel interaction conditions (none, slight, moderate, and severe) were tested in separate blocks and the test block order was randomized across subjects. During each test block, a contour was randomly selected (without replacement) from among the 54 stimuli (9 contours \times 3 semitone spacings \times 2 repeats) and presented to the subject, who responded by clicking on one of the nine response boxes shown onscreen. Subjects were allowed to repeat each stimulus up to three times. No preview or trial-by-trial feedback was provided. A minimum of two test blocks were tested for each channel interaction condition; if the difference in performance was greater than 10%, a third run was performed. The scores for all trials were then averaged together.

3. Results

All subjects scored above 90% correct with the unprocessed stimuli. Figure 3 shows mean MCI performance as a function of semitone spacing (left panel) and degree of channel interaction averaged across semitone conditions (right panel). MCI performance monotonically worsened as the amount of channel interaction was increased. A two-way repeated-measures analysis of variance showed significant main effects for channel interaction [$F_{(3,90)} = 149.4$, $p < 0.001$] and semitone spacing [$F_{(2,90)} = 203.1$, $p < 0.001$], as well as a significant interaction [$F_{(6,90)} = 15.6$, $p < 0.001$]. *Post hoc* Bonferroni comparisons revealed significant differences between all channel interaction conditions ($p < 0.001$ in all cases) and between all semitone spacing conditions ($p < 0.001$ in all cases).

4. Discussion

As hypothesized, the present CI simulation results show that channel interaction can negatively affect melodic pitch perception. As illustrated in Fig. 2, spectral envelope cues are weakened by CI signal processing and further weakened by channel interaction. As such, increasing the number of channels may not sufficiently enhance spectral contrasts between notes. Most CI signal processing strategies use monopolar stimulation, which results in broader activation and greater channel interaction than with current-focused stimulation (e.g., tripolar stimulation) (Bierer, 2007). The present data suggest that reducing channel interaction may be as important a goal as increasing the number of stimulation sites.

In this study, the amount of channel interaction was constant across subjects, and constant across channels within each condition. In the real CI case, channel interaction may vary greatly across CI users, and across electrode location within CI users. Interestingly, mean performance with real CI users for the exactly the same task and stimuli was 61% correct (Zhu *et al.*, 2011; dashed line in Fig. 3), and was most comparable to mean CI simulation performance with slight channel interaction in the present study. Note that the present NH subjects had no prior experience listening to vocoded sounds, compared with years of experience with electric stimulation for real CI users. With more experience, NH performance would probably improve, but the general trend across conditions would most likely remain. It is possible that the effects of channel interaction observed in this study may explain some of the larger variability observed in Zhu *et al.* (2011), with some CI users experiencing moderate-to-severe channel interaction and others experiencing very little. Of course, many other factors can contribute to CI users' melodic pitch perception (e.g., acoustic frequency allocation, electrode location, pattern of nerve survival, experience, etc.).

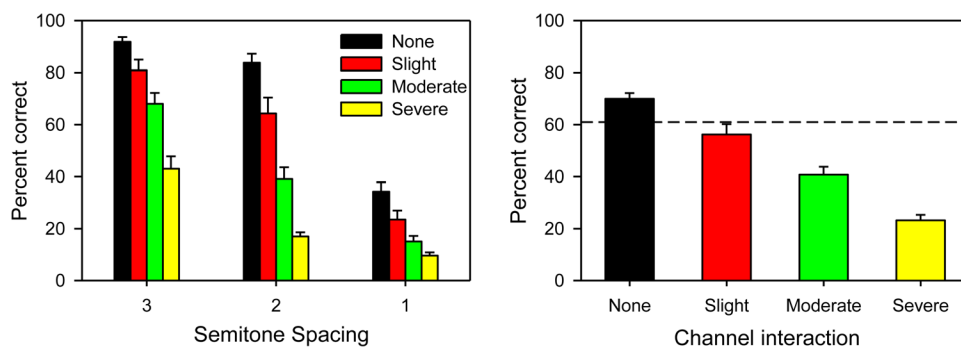


Fig. 3. (Color online) Mean MCI performance with the CI simulation as a function of semitone spacing (left panel) or the degree of channel interaction (right panel; performance averaged across semitone spacing conditions). The error bars indicate the standard error. The dashed line in the right-hand panel shows mean CI performance from Zhu *et al.* (2011).

The number of spectral channels has been shown to limit CI performance in difficult listening situations (Friesen *et al.*, 2001; Kong *et al.*, 2004; Luo *et al.*, 2007). Data from the present study and from Fu and Nogaki (2005) suggest that channel interaction may also limit CI performance where perception of pitch cues may be beneficial. In dynamic noise, pitch cues may help to stream a talker's voice and segregate target speech from dynamic noise or a competing talker; when channel interaction is increased, pitch cues may become less salient and segregation more difficult. In previous CI simulation studies, the amount of channel interaction did not necessarily increase as the number of channels increased, as would happen in the real CI case. In these studies, CI simulation performance improved as the number of channels increased, whereas real CI performance peaked at six to ten channels, presumably due to channel interaction. The present MCI data suggest that pitch perception worsens with increasing channel interaction. As the CI does not provide strong pitch cues, channel interaction may further weaken already poor pitch perception.

The present data imply that increasing the number of stimulation sites, whether with more electrodes or virtual channels, may not be sufficient to provide adequate pitch cues. The present study essentially simulated a discrete neural population with 16 fixed channels; the channel interaction conditions simulated increased current spread across these locations. Because the locations were fixed, the change in the spectral envelope was the dominant cue for melodic pitch. As seen in Fig. 2, as the channel interaction increased, the variance in the spectral envelope was reduced, making pitch perception more difficult. Although increasing the number of stimulation sites would seem to increase the spectral resolution, the present suggest a need to also limit channel interaction. Current focusing (Landsberger *et al.*, 2012) or optical stimulation (Izzo *et al.*, 2007) may help to reduce channel interaction in CIs and other auditory neuroprostheses. In Landsberger *et al.* (2012), the perceptual quality of electric stimulation (e.g., clarity, purity, fullness, etc.) improved with current focusing and was correlated with reduced spread of excitation. Reducing the spread of excitation might reduce channel interaction, which, according to present study, would improve melodic pitch perception. There may be an optimal tradeoff between the number of channels and the degree of channel interaction as spread of excitation also occurs to some extent in acoustic hearing.

Acknowledgments

The authors thank the subjects for their participation. This work was supported by the USC Provost Fellowship, the USC Hearing and Communication Neuroscience Program, NIDCD R01-DC004993, and by the Paul Veneklasen Research Foundation.

References and links

- Bierer, J. A. (2007). "Threshold and channel interaction in cochlear implant users: evaluation of the tripolar electrode configuration," *J. Acoust. Soc. Am.* **121**, 1642–1653.
- Bingabr, M., Espinoza-Varas, B., and Loizou, P. C. (2008). "Simulating the effect of spread of excitation in cochlear implants," *Hear. Res.* **241**, 73–79.
- Dorman, M. F., Loizou, P. C., and Rainey, D. (1997). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.* **102**, 2403–2411.
- Friesen, L. M., Shannon, R. V., Baskent, D., and 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.
- Fu, Q.-J., and Nogaki, G. (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," *J. Assoc. Res. Otolaryngol.* **6**, 19–27.
- Galvin, J. J., III, Fu, Q.-J., and Nogaki, G. (2007). "Melodic contour identification by cochlear implant listeners," *Ear Hear.* **28**, 302–319.
- Galvin, J. J., III, Fu, Q.-J., and Oba, S. (2008). "Effect of instrument timbre on melodic contour identification by cochlear implant users," *J. Acoust. Soc. Am.* **124**, EL189–EL195.

- Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J., Witt, S., and Stordahl, J. (2002). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," *Coch. Imp. Inter.* **3**, 29–53.
- Greenwood, D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Izzo, A. D., Suh, E., Pathria, J., Walsh, J. T., Whitlon, D. S., and Richter, C.-P. (2007). "Selectivity of neural stimulation in the auditory system: A comparison of optic and electric stimuli," *J. Biomed. Opt.* **12**, 021008.
- Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (2004). "Music perception with temporal cues in acoustic and electric hearing," *Ear Hear.* **25**, 173–185.
- Landsberger, D. M., Padilla, M., and Srinivasan, A. G. (2012). "Reducing current spread using current focusing in cochlear implant users," *Hear. Res.* **284**, 16–24.
- Luo, X., Fu, Q.-J., and Galvin, J. (2007). "Vocal emotion recognition by normal-hearing listeners and cochlear implant users," *Trends Amplif.* **11**, 301–315.
- McDermott, H. J. (2004). "Music perception with cochlear implants: A review," *Trends Amplif.* **8**, 49–82.
- Zhu, M., Chen, B., Galvin, J., and Fu, Q.-J. (2011). "Influence of pitch, timbre and timing cues on melodic contour identification with a competing masker," *J. Acoust. Soc. Am.* **130**, 3562–3565.