Contour identification with pitch and loudness cues using cochlear implants

Xin Luo,^{a)} Megan E. Masterson, and Ching-Chih Wu

*Department of Speech, Language, and Hearing Sciences, Purdue University,
500 Oval Drive, West Lafayette, Indiana 47907
luo5@purdue.edu, memaster@purdue.edu, wu94@purdue.edu*

Abstract: Different from speech, pitch and loudness cues may or may not co-vary in music. Cochlear implant (CI) users with poor pitch perception may use loudness contour cues more than normal-hearing (NH) listeners. Contour identification was tested in CI users and NH listeners; the five-note contours contained either pitch cues alone, loudness cues alone, or both. Results showed that NH listeners' contour identification was better with pitch cues than with loudness cues; CI users performed similarly with either cues. When pitch and loudness cues were co-varied, CI performance significantly improved, suggesting that CI users were able to integrate the two cues.

© 2013 Acoustical Society of America

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

Date Received: September 11, 2013 **Date Accepted:** November 8, 2013

1. Introduction

Normal-hearing (NH) listeners' melody recognition is primarily based on relative pitch changes across notes rather than exact pitches of individual notes; pitch contours and intervals (i.e., the directions and sizes of pitch changes, respectively) are both critical for melody recognition (e.g., [Dowling and Fujitani, 1970](#)). Melody recognition remains a challenge for most cochlear implant (CI) users, due to many technological and perceptual deficits with CIs (e.g., the small number of electrodes, the large current spread of electric stimulation, and the degraded temporal pitch cues above 300 Hz). For example, CI users have been shown to have significantly poorer melodic contour and interval perception than NH listeners (e.g., [Galvin *et al.*, 2007](#); [Gfeller *et al.*, 2007](#)).

Although musical melodies are typically created with pitch changes across notes, analogues of musical melodies are possible in other auditory dimensions such as loudness ([McDermott *et al.*, 2008](#)). NH listeners can reliably identify contours across the pitch and loudness dimensions, as long as the directions and relative sizes of the changes across notes are similar in both dimensions (i.e., higher pitches mapped to louder sounds and vice versa). Such loudness cues can also be used to recognize familiar melodies, although the recognition accuracy is poorer than that with pitch cues ([McDermott *et al.*, 2008](#)). Using dichotic listening tests, [Neuhoff *et al.* \(1999\)](#) showed that pitch and loudness contour cues may interact with each other centrally.

This study tested possible interactions between pitch and loudness cues in a musically relevant contour identification task in CI users. As in NH listeners, pitch and loudness contour cues may have common central representations in CI users. However, degraded peripheral inputs with CIs may adversely affect contour perception both within and across the dimensions. [Cousineau *et al.* \(2010\)](#) found that with equally discriminable sequence elements, NH listeners had better pitch than loudness contour discrimination, while CI users had similar pitch or loudness contour discrimination. Compared to NH listeners, CI users exhibit deficits in pitch but not loudness contour processing. As such, CI users may use loudness cues more strongly than NH listeners

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

in contour identification tests where pitch and/or loudness cues are varied. Here, contour identification was tested in CI users and NH listeners, using pitch cues alone, loudness cues alone, or consistent pitch and loudness cues (i.e., both cues varied in the same directions). Different semitone spacing and intensity changes between successive notes were used to vary the perceptual salience of pitch and loudness cues, respectively, as the integration of both cues may depend on their relative salience. Previously, we have found that CI users' Mandarin tone recognition was enhanced when the amplitude envelope was modified to follow the pitch contour of lexical tone (Luo and Fu, 2004). In this study, CI users' contour identification is hypothesized to be better when pitch and loudness cues are consistent.

2. Methods

2.1 Subjects

Six college-aged female NH listeners served as the control. Ten post-lingually deafened adult CI users were tested with clinical processors using the clinically assigned sensitivity and volume settings. Table 1 lists CI subject demographics. All CI subjects had at least one year of CI use. Bilateral CI users S2, S7, and S8 were tested with their first implant only. Bimodal CI users S4, S5, and S10 were tested without the hearing aid. The remaining CI subjects were unilateral CI users. Only one NH and three CI subjects (S2, S4, and S10) have had several years of musical training and group activity from elementary school to college. This study was approved by the IRB committee of Purdue University. All subjects gave informed consent and were paid for their participation.

2.2 Determining the usable intensity range

Loudness ratings of broadband noise (Allen *et al.*, 1990) were used to determine the usable intensity range for loudness contours. A 250-ms white noise was filtered by a fourth-order Butterworth filter with cutoff frequencies at 250 and 1750 Hz and then added with 10-ms onset and offset raised cosine ramps. The broadband noise was presented in a double-walled sound-treated booth via a single loudspeaker at levels ranging from 40 to 80 dB sound pressure level (SPL) in 5-dB steps, as measured at the ear level of participants. Each subject rated the loudness of the noise using a six-point scale that ranged from "no sound" to "too loud." Each intensity level was previewed once and then tested three times in a random order. Figure 1(a) shows the mean intensity levels for different loudness ratings for NH listeners and CI users. Both groups exhibited monotonic loudness growth and on average, CI users needed significantly higher SPLs than did NH listeners for the "comfortable" and "soft" loudness ratings ($p < 0.001$), as revealed by a two-way analysis of variance (ANOVA). Figure 1(b) shows that the loudness growth functions largely overlap for Advanced Bionics and Cochlear CI users, even though the input dynamic range (IDR) is smaller in the

Table 1. CI subject demographics.

Subject	Age	Gender	Etiology	Device	Strategy	Years with CI
S1	43	Female	Sudden hearing loss	HiRes90K	HiRes-120	7
S2	29	Male	Otosclerosis	Nucleus 5	ACE	1
S3	48	Male	Sudden hearing loss	Clarion 1.0	CIS	16
S4	83	Female	Sudden hearing loss	HiRes90K	HiRes-120	3
S5	69	Male	Noise exposure	Nucleus 5	ACE	2
S6	63	Female	Gestational hypertension	HiRes90K	HiRes-P	6
S7	61	Female	Nerve death	Nucleus 24	ACE	10
S8	66	Male	Noise and diabetes	Freedom	ACE	5
S9	79	Female	Hereditary deafness	Freedom	ACE	5
S10	54	Female	Unknown	Freedom	ACE	5

Fig. 1. Loudness growth as a function of noise intensity level (a) for NH listeners and CI users and (b) for Advanced Bionics and Cochlear CI users. Symbols show the mean intensity level in dB SPL, and error bars show the standard deviation across subjects for the corresponding loudness rating. For clarity of illustration, error bars are shown in only one direction.

Cochlear devices (30–40 dB) than in the Advanced Bionics devices (60–80 dB). Based on the loudness-rating results, two intensity ranges (45–75 or 60–75 dB SPL) were used to create loudness contours of different perceptual salience. These ranges supported good loudness growth for both NH and CI subjects.

2.3 Contour identification with pitch and loudness cues

Similar to Galvin *et al.* (2007), each pitch contour consisted of a sequence of five notes with nine possible contour patterns (rising, rising-flat, rising-falling, flat-rising, flat, flat-falling, falling-rising, falling-flat, and falling). Each note was a harmonic complex tone with fundamental frequency (F_0) presented at 0 dB, $2 \times F_0$ at -3 dB, and $3 \times F_0$ at -6 dB. The lowest note of the nine pitch contours, defined as the root note, was varied to be A3 (220 Hz), A4 (440 Hz), or A5 (880 Hz) to test contour identification in different F_0 ranges. The semitone spacing between successive notes was varied (1, 3, or 5 semitones) to adjust the perceptual salience of pitch contours; an n -semitone spacing has a frequency ratio of $2^{(n/12)}$. Each note in the pitch contours was presented at 70 dB SPL. There were a total of 81 pitch contours (3 root notes $\times 3$ semitone spacing $\times 9$ contour patterns).

Loudness contours with the same nine patterns were generated by varying the intensity levels of a sequence of five broadband noise bursts. The same noise bursts in the loudness-rating test were used instead of harmonic complex tones to avoid interfering pitch cues. The intensity change between successive noise bursts was 7.5 dB for the 30-dB intensity range (45–75 dB SPL) and 3.75 dB for the 15-dB range (60–75 dB SPL). There were a total of 18 loudness contours (2 intensity ranges $\times 9$ contour patterns).

Pitch-loudness contours were generated by varying the intensity levels of the five notes in the pitch contours in the same directions as the pitch changes (i.e., increases in F_0 were combined with increases in intensity and vice versa). The intensity changes in the pitch-loudness contours were the same as those in the loudness contours. The overall F_0 range of the nine pitch contours (regardless of the root note and semitone spacing) was mapped to the 30- or 15-dB intensity range. There were a total of 162 pitch-loudness contours (3 root notes $\times 3$ semitone spacing $\times 2$ intensity ranges $\times 9$ contour patterns).

For all contours, individual notes or noise bursts were 250 ms with 10-ms onset and offset raised cosine ramps. There was a 50-ms silence gap between successive notes.

Contour identification was tested using a nine-alternative, forced-choice task. The testing order of cue conditions was randomized within and across subjects. Before testing, subjects previewed the 18 loudness contours, a subset of 18 pitch contours (A3 root note with 1- or 3-semitone spacing), and a subset of 18 pitch-loudness contours (A3 root note with 1- or 3-semitone spacing and the 15-dB intensity range). During testing, a contour was randomly selected from the stimulus set (without replacement) and presented to the subject, who responded by clicking on one of the nine response buttons that displayed the nine possible contours. No feedback was provided. Loudness contours were tested twice due to the small number, while pitch and pitch-loudness contours were tested only once. The percent correct scores were recorded for each subject in each condition.

3. Results

Figure 2(a) shows overall contour identification results with either pitch cues alone, loudness cues alone, or consistent pitch and loudness cues for NH listeners and CI

Fig. 2. (a) Mean contour identification for different cue conditions, and (b) as a function of root note, (c) semitone spacing, and (d) intensity range with either pitch cues alone, loudness cues alone, or consistent pitch and loudness cues. The white boxes show NH performance and the gray boxes show CI performance. The lines within the boxes show the median and the box boundaries show the 75th and 25th percentiles. The error bars show the 90th and 10th percentiles and the dots show outliers. The dashed horizontal lines indicate chance performance level (11.11% correct).

users. A mixed-design ANOVA showed that contour identification was significantly different between subject groups ($F_{1,14} = 13.92$, $p = 0.002$) and across cue conditions ($F_{2,28} = 13.78$, $p < 0.001$). There was no significant interaction between the two factors ($F_{2,28} = 2.64$, $p = 0.09$). The NH and CI data were then separately analyzed using one-way repeated-measures (RM) ANOVAs and *post hoc* Bonferroni *t*-tests. NH performance with consistent pitch and loudness cues was as perfect as that with pitch cues alone, and both were significantly better (18% higher on average) than that with loudness cues alone ($p < 0.01$). CI users performed similarly with either pitch or loudness cues alone. Their performance with consistent pitch and loudness cues was significantly better (13% higher on average) than that with loudness cues alone ($p = 0.02$) and was marginally better (10% higher on average) than that with pitch cues alone ($p = 0.06$).

Figure 2(b) shows NH and CI results with either pitch cues alone or consistent pitch and loudness cues as a function of root note. NH performance was near perfect with all three root notes in both cue conditions. For CI users, a two-way RM ANOVA showed a significant effect for cue condition ($F_{1,9} = 13.66$, $p = 0.01$) but not for root note ($F_{2,18} = 1.16$, $p = 0.34$). The two factors did not interact with each other ($F_{2,18} = 0.02$, $p = 0.98$), suggesting that adding loudness cues improved CI performance with all three root notes (by about 10% on average).

Figure 2(c) shows NH and CI results with either pitch cues alone or consistent pitch and loudness cues as a function of semitone spacing. Again, NH performance was near perfect with all three semitone spacing in both cue conditions. A two-way RM ANOVA on the CI data showed a significant effect for both cue condition ($F_{1,9} = 13.05$, $p = 0.01$) and semitone spacing ($F_{2,18} = 34.55$, $p < 0.001$); the two factors significantly interacted with each other ($F_{2,18} = 15.52$, $p < 0.001$). *Post hoc* Bonferroni *t*-tests showed that adding loudness cues significantly improved CI performance with the 1-semitone spacing by 22% on average ($p < 0.001$), but not with the 3- or 5-semitone spacing. In the pitch-only condition, CI performance with the 3- or 5-semitone spacing was significantly better (24% or 30% higher on average, respectively) than that with the 1-semitone spacing ($p < 0.001$). In the pitch-loudness condition, CI performance with the 5-semitone spacing remained significantly better (11% higher on average) than that with the 1-semitone spacing ($p = 0.005$).

Figure 2(d) shows NH and CI results with either loudness cues alone or consistent pitch and loudness cues for the two intensity ranges. A two-way RM ANOVA on the NH data showed a significant effect for cue condition ($F_{1,5} = 16.03$, $p = 0.01$) but not for intensity range ($F_{1,5} = 5.36$, $p = 0.07$); there was a significant interaction between the two factors ($F_{1,5} = 7.68$, $p = 0.04$). *Post hoc* Bonferroni *t*-tests showed that NH performance with the 30-dB intensity range was significantly better (13% higher on average) than that with the 15-dB intensity range, but only in the loudness-only condition ($p = 0.005$). Combining pitch and loudness cues significantly improved NH performance with either the 15- or 30-dB intensity range by 26% or 12% on average, respectively ($p < 0.05$). A two-way RM ANOVA on the CI data also showed a significant effect for cue condition ($F_{1,9} = 11.03$, $p = 0.01$) but not for intensity range ($F_{1,9} = 1.28$, $p = 0.29$). The two factors did not interact with each other ($F_{1,9} = 1.30$, $p = 0.28$), suggesting that adding pitch cues significantly improved CI performance with either the 15- or 30-dB intensity range by 19% or 8% on average, respectively.

4. Discussion

In this study, NH listeners' contour identification was better with pitch cues than with loudness cues. CI users performed similarly with either cues alone. The hypothesis that consistent pitch and loudness cues may improve contour identification with CIs was true, but only for pitch contours with the 1-semitone spacing.

Pitch contour identification was better with the present CI subjects (67% correct) than with those in Galvin *et al.* (2007; 53% correct). Newer generations of CI devices and processing strategies were used in this study than in Galvin *et al.* (2007), which may have contributed to performance difference. However, there was no clear

effect of different CI devices and processing strategies in either study. CI users' pitch contour identification was similar with different root notes, possibly due to the tradeoff between temporal and place pitch cues. Temporal pitch cues may be available only for contours with the lowest root note A3, while place coding may be better for contours with higher root notes A4 or A5. In current CIs, the spectral resolution is generally greater and the acoustic frequency-to-electrode place mismatch is less severe for contours with higher F0s than with lower F0s (Singh *et al.*, 2009). CI users had significantly poorer pitch contour identification with smaller semitone spacing, especially when successive notes were separated by only 1 semitone. This was in contrast with the near-perfect NH performance even with the 1-semitone spacing and reflected the limited pitch discrimination ability with CIs.

In line with the contour discrimination results of Cousineau *et al.* (2010), NH listeners' loudness contour identification was poorer than pitch contour identification; CI users performed similarly with either cues alone. Unlike Cousineau *et al.* (2010), we did not equalize the discriminability of pitch and loudness changes between successive notes in the contours for individual subjects. Given the mean pitch and loudness discrimination thresholds in Cousineau *et al.* (2010), the present NH listeners may have perceived more pitch steps in the semitone spacing than loudness steps in the intensity changes between successive notes. In contrast, the present CI users may have perceived similar numbers of pitch and loudness steps between successive notes, which were both less than the step numbers perceived by NH listeners.

NH listeners' loudness contour identification was significantly better with the 30-dB intensity range than with the 15-dB range, because the intensity changes and the number of perceived loudness steps between successive notes were doubled with the 30-dB range. However, not all CI users benefitted from the larger intensity range. Some CI users' loudness contour identification may have been adversely affected by the low SPLs of some notes in the 45–75 dB intensity range.

NH listeners' near-perfect pitch contour identification left little room for further improvement in contour identification with consistent pitch and loudness cues. On the other hand, CI users had significantly better contour identification with consistent pitch and loudness cues than with either cues alone. McDermott *et al.* (2008) found that, with the relative sizes of intervals preserved, pitch and loudness changes in the same directions elicited common contour representations that were recognized across the two dimensions by NH listeners. Our results suggest that when pitch and loudness changed together, their common contour representations may be integrated to enhance contour identification with CIs. The central mechanism for cue integration (e.g., Neuhoff *et al.*, 1999) also worked in CI users for degraded peripheral inputs. Better contour identification was found with only the smallest 1-semitone spacing, suggesting that CI users relied more heavily on loudness contour cues when pitch contour cues were weaker. When pitch cues were more salient with 3- and 5-semitone spacing, loudness cues did not contribute, although ceiling effects were only found for CI subjects S2 and S3. Even with consistent pitch and loudness cues, CI users' contour identification remained much poorer than that of NH listeners. Compared with younger NH listeners, older CI users may have had both CI- and age-related deficits in discrimination and contour processing of pitch and loudness changes.

Consistent changes in intensity with F0 may affect the pitch percepts of harmonic complexes and lead to more salient pitch contours. The level-induced pitch changes of harmonic complexes were examined in the first five CI subjects using a pitch-matching test. A 250-ms pure tone was presented at 70 dB SPL after a 250-ms 3-harmonic complex tone A4 (440 Hz) presented at 45, 60, or 75 dB SPL in different conditions. The frequency of pure tone was adjusted until its pitch matched that of the complex tone. The pitch-matching errors from 440 Hz suggest that pitch of the complex tone increased by seven semitones when its level increased from 45 to 60 dB SPL but saturated for levels from 60 to 75 dB SPL. As the level decreased to 45 dB SPL, higher harmonics of the complex tone may have become inaudible first because its

spectrum had a high-frequency roll-off. The corresponding electrodes may have been deactivated and the centroid of excitation may have been shifted apically, leading to lower pitch percepts. The pitch decreases at only low levels may have partially contributed to CI users' better contour identification when F0 and intensity varied in the same directions.

In summary, CI subjects were able to use both pitch and loudness cues in contour identification, especially when pitch cues were perceptually weak. Pitch and loudness cues may or may not co-vary with each other in real-life musical melodies. When they do, contour patterns or even melodies may be better perceived with CIs, as suggested by the present results. Future studies should consider situations where the two cues conflict with each other (e.g., pitch contours with intensity roving) to see if contour identification would be adversely affected. A practical implication of the current findings for CI signal processing strategies is that pitch and loudness cues may be artificially co-varied to ease music perception for CI users. This hypothesis may be first tested by measuring melody recognition with CIs in different cue conditions.

Acknowledgments

We are grateful to all subjects for their participation in this study. We thank John Galvin for editorial assistance. Research was supported in part by NIH (R21-DC011844).

References and Links

- Allen, J. B., Hall, J. L., and Jeng, P. (1990). "Loudness growth in 1/2-octave bands (LGOB)—A procedure for the assessment of loudness," *J. Acoust. Soc. Am.* **88**(2), 745–753.
- Cousineau, M., Demany, L., Meyer, B., and Pressnitzer, D. (2010). "What breaks a melody: Perceiving F0 and intensity sequences with a cochlear implant," *Hear Res.* **269**(1–2), 34–41.
- Dowling, W. J., and Fujitani, D. S. (1970). "Contour, interval, and pitch recognition in memory for melodies," *J. Acoust. Soc. Am.* **49**(2), 524–531.
- Galvin, J. J., Fu, Q.-J., and Nogaki, G. (2007). "Melodic contour identification by cochlear implant listeners," *Ear Hear.* **28**(3), 302–319.
- Gfeller, K., Turner, C., Oleson, J., Zhang, X. Y., Gantz, B., Froman, R., and Olszewski, C. (2007). "Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise," *Ear Hear.* **28**(3), 412–423.
- Luo, X., and Fu, Q.-J. (2004). "Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants," *J. Acoust. Soc. Am.* **116**(6), 3659–3667.
- McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2008). "Is relative pitch specific to pitch?" *Psychol. Sci.* **19**(12), 1263–1271.
- Neuhoff, J. G., McBeath, M. K., and Wanzie, W. C. (1999). "Dynamic frequency change influences loudness perception: A central, analytic process," *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1050–1059.
- Singh, S., Kong, Y.-Y., and Zeng, F.-G. (2009). "Cochlear implant melody recognition as a function of melody frequency range, harmonicity, and number of electrodes," *Ear Hear.* **30**(2), 160–168.