Melodic interval perception by normal-hearing listeners and cochlear implant users

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

(Received 20 February 2014; revised 5 August 2014; accepted 12 August 2014)

The perception of melodic intervals (sequential pitch differences) is essential to music perception. This study tested melodic interval perception in normal-hearing (NH) listeners and cochlear implant (CI) users. Melodic interval ranking was tested using an adaptive procedure. CI users had slightly higher interval ranking thresholds than NH listeners. Both groups' interval ranking thresholds, although not affected by root note, significantly increased with standard interval size and were higher for descending intervals than for ascending intervals. The pitch direction effect may be due to a procedural artifact or a difference in central processing. In another test, familiar melodies were played with all the intervals scaled by a single factor. Subjects rated how in tune the melodies were and adjusted the scaling factor until the melodies sounded the most in tune. CI users had lower final interval ratings and less change in interval rating as a function of scaling factor than NH listeners. For CI users, the root-mean-square error of the final scaling factors and the width of the interval rating function were significantly correlated with the average ranking threshold for ascending rather than descending intervals, suggesting that CI users may have focused on ascending intervals when rating and adjusting the melodies. © 2014 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4894738>]

PACS number(s): 43.66.Ts, 43.66.Hg, 43.75.Cd [DB]

Pages: 1831–1844

I. INTRODUCTION

In music, the sizes of pitch change between sequentially presented notes (i.e., melodic intervals) are expressed in semitones, a logarithmic measure of frequency ratio. The perceived interval sizes are roughly the same for equal distances along such a log-frequency axis (e.g., [Attneave and Olson, 1971](#)) and linearly increase with frequency change in semitones (e.g., [Russo and Thompson, 2005](#)). Melodic intervals and contours (i.e., the directions of pitch change) are two components of relative pitch that contribute to the perception, appreciation, and memory of melodies (e.g., [Dowling and Fujitani, 1970](#); [Dowling, 1978](#); [Cuddy and Cohen, 1976](#)). Music composed from different intervals and scales (e.g., the major or minor scale) may indicate different music types and evoke different moods (e.g., [Hevner, 1935](#); [Maher, 1980](#)).

Because melodic interval perception is so important to music perception, it has been extensively studied in normal-hearing (NH) listeners using different listening tasks. For example, the perception of melodic intervals according to Western music labels such as minor second, major second, and so on has been shown to be more categorical and less affected by context in musicians than in non-musicians (e.g., [Siegel and Siegel, 1977](#); [Burns and Ward, 1978](#); [Zatorre and Halpern, 1979](#)). In some studies (e.g., [Ward, 1954](#); [Rakowski, 1990](#)), musicians were asked to produce certain melodic intervals by adjusting a variable tone with reference to a preceding fixed tone. For example, the production

results of musicians for an octave (or 12 semitones) had an average standard deviation of 10 cents (a cent is 1/100 of a semitone) ([Ward, 1954](#)). [Rakowski \(1990\)](#) found that the standard deviation of the production results for a certain interval was correlated with the percentage of errors in recognizing that interval. Ranking thresholds have been measured for only ascending intervals using the method of constant stimuli (e.g., [Burns and Ward, 1978](#); [Burns and Campbell, 1994](#); [Zarate et al., 2012](#)) or adaptive procedures ([Burns and Ward, 1978](#); [McDermott et al., 2010](#)). On average, ascending interval ranking thresholds were below 1 semitone (i.e., the smallest tonal unit in Western music) for musicians, but above 1 semitone for non-musicians. However, the effects of root note (i.e., the lower note of an interval), pitch change direction, and standard interval size on interval ranking have not been fully explored. These factors may interact with each other, as shown in the subjective size estimation of intervals ([Russo and Thompson, 2005](#)). In summary, these previous studies with NH listeners clearly showed an effect of musical experience on melodic interval perception, with musicians performing better than non-musicians. Also, identification and production of isolated melodic intervals may not be suitable for testing non-musicians with limited musical experience.

Profoundly deaf people using cochlear implants (CIs) have great difficulty in music perception, in contrast with their relatively good speech recognition in quiet (e.g., [Gfeller et al., 2002](#); [Kong et al., 2004](#); [Cooper et al., 2008](#); [Looi et al., 2008](#); [Nimmons et al., 2008](#)). Current CI processors only encode slowly varying temporal envelopes in a small number of spectral channels, which cannot support accurate pitch perception in music (e.g., [Galvin et al., 2007](#); [Cooper et al., 2008](#)). As

^{a)}Author to whom correspondence should be addressed. Electronic mail: luo5@purdue.edu

such, CI users strongly rely on rhythm cues to recognize familiar melodies (Kong *et al.*, 2004). Given the importance of interval processing to music perception, there is a great need for research to identify factors that may affect melodic interval perception of CI users.

A few early studies (Pijl and Schwarz, 1995; Pijl, 1997; McDermott and McKay, 1997) have specifically tested melodic interval perception with CIs using direct electric stimulation (i.e., bypassing clinical speech processors). Pijl and Schwarz (1995) and Pijl (1997) asked three CI users to rate the melodic intervals between different pulse rates on a single electrode as “in tune,” “sharp,” or “flat,” relative to their memory of a specific interval in a familiar melody (e.g., the perfect fifth or 7-semitone interval between the second and third notes of “Twinkle, Twinkle, Little Star”). The interval rating functions were analyzed to estimate the point of subjective equality (PSE) or the ratio of pulse rates that was perceived to match the target interval. CI users’ PSEs were within one semitone of the target interval, suggesting that temporal cues from low pulse rates (<500 Hz) may be sufficient for melodic interval perception at least for some CI users. In contrast, there were greater errors in interval perception when CI users listened to acoustic stimuli via their clinical processors with the SPEAK strategy. McDermott and McKay (1997) varied the pulse rate or amplitude modulation frequency to encode temporal pitch cues and used the different electrodes to encode place pitch cues. A CI user qualified as a musical instrument tuner before implantation was asked to identify and produce Western music intervals based on temporal cues alone, place cues alone, or both. It was found that temporal pitch cues supported near-normal interval perception only for frequencies <500 Hz. Place pitch cues also conveyed interval information and dominated pitch perception when both pulse rate and stimulation place changed. However, these interval perception tasks may not be applicable to general CI population who has limited knowledge of musical terminology or experience with melodic intervals, making it difficult for them to isolate an interval from the rhythm and pitch contour of a familiar melody. Although direct CI stimulation with variable pulse rates has been a preferred method to study the temporal mechanisms of interval perception, its results do not represent those with clinical CI processors, which stimulate multiple electrodes with amplitude-modulated, high-rate pulse trains to use both place and temporal cues (although not as well controlled as in direct electric stimulation; see the example outputs of CI processors in Sec. II B).

A number of test batteries (e.g., Cooper *et al.*, 2008; Kang *et al.*, 2009) have recently been used or developed to evaluate CI users’ perception of different music elements (e.g., melody, timbre, rhythm, meter, contour, interval, and scale) using clinical processors. In the interval test of the Montreal Battery for Evaluation of Amusia (MBEA; Peretz *et al.*, 2003), two melodies were played in each of 30 trials and subjects would judge whether the two melodies were the same or different. In half of the trials, a note in the comparison melody was altered to change its related intervals while preserving the pitch contour of the original melody. Cooper *et al.* (2008) found that CI users performed at chance level

(50% correct) in this test using clinical processors. However, the overall percent correct score did not reveal enough details of a CI user’s interval perception ability (e.g., how much interval change can be perceived and what factors may affect interval perception). Wang *et al.* (2011) measured the interval discrimination thresholds of 16 CI users in the context of familiar melodies using an adaptive procedure. The fifth note of “Twinkle, Twinkle, Little Star” and the sixth note of “Happy Birthday” were changed adaptively in a same/different task similar to that in the MBEA test. For the two tested notes, CI users had a wide range of interval discrimination thresholds from 0.2 to 20 semitones with a mean of 5.7 semitones. However, the note that was changed from trial to trial was always the tested note and the other notes in the melody were fixed throughout the test. Listeners may thus have focused on the tested note alone and performed pitch rather than interval discrimination. Swanson (2008) also developed a modified melodies test to study interval perception in the context of familiar melodies. In the stretched distortion condition, all the intervals in a familiar melody (e.g., “Twinkle, Twinkle, Little Star”) were scaled by a single factor. Both the modified and original versions of the melody were played in each trial (in random order) and listeners were asked to identify the original or correct version. The results showed that NH listeners already obtained perfect scores (100% correct) for scaling factors close to 1 (e.g., 0.9 and 1.1), while CI users still performed at chance level (50% correct) even for scaling factors as large as 2 or as small as 0.5. Thus, it is difficult to find scaling factors for the test to avoid both ceiling effects in NH performance and floor effects in CI performance (measured in percent correct).

In this study, two tests of melodic interval perception were adapted from those of previous studies to facilitate the testing of NH listeners and CI users with limited musical knowledge or experience. The goal is to better understand the deficits in melodic interval perception of CI users with the latest implant technologies. The first test (melodic interval ranking test) was similar to that of McDermott *et al.* (2010), except that the root note, pitch direction, and standard interval size were systematically varied to investigate their effects on the fidelity of interval perception in NH listeners and CI users. To ensure that subjects had to compare the sizes of the two intervals in each trial to perform the ranking task, frequency roving was applied to the root note of each interval. Basic pitch ranking was also tested as a baseline measure of pitch sensitivity. It was hypothesized that CI users would have significantly poorer basic pitch ranking and melodic interval ranking thresholds than NH listeners. The root note may affect CI results (but not NH results) because different amounts of place and temporal pitch cues may be available for different root notes via the CI processors (see Sec. II B for details). Based on the interval size estimation data of Russo and Thompson (2005) and the distribution of different intervals in Western music (Vos and Troost, 1989), interval ranking thresholds were also expected to vary with interval size and pitch direction. In another test, NH listeners and CI users were asked to rate and adjust (or correct) the different intervals in the initial

phrases of familiar melodies (rather than isolated intervals; e.g., Pijl, 1997; McDermott and McKay, 1997). Subjects with limited musical experience may better remember intervals in the context of familiar melodies. The existence of melodic context may also reduce the variation in interval adjustment (Rakowski, 1990). As in the modified melodies test (Swanson, 2008), all the intervals in a melody were scaled by a single factor to simplify the interval adjustment, although CI users often wanted to compress or expand different intervals to different degrees. Our hypothesis was that compared to NH listeners, CI users would rate the melodies as more out of tune with lower final interval ratings and may prefer physically larger intervals created by larger final scaling factors, due to their poorer pitch perception. Being less sensitive to interval expansions or compressions, CI users may reduce their interval ratings more slowly with increasing or decreasing interval scaling factors (as indicated by a greater width of the interval rating function) and may have more errors in the final scaling factors for different melodies and test conditions. Finally, these measures of interval rating and adjustment (i.e., the width of the interval rating function and the root-mean-square error of the final scaling factors) might be correlated with the melodic interval ranking thresholds across subjects, similar to the correlation between interval production and perception reported by Rakowski (1990).

II. METHODS

A. Subjects

The NH control group had six college-aged female NH listeners (NH1-NH6 aged from 21 to 27 years old with a mean of 24 years old), whose pure-tone thresholds were below 25 dB hearing level (HL) at octave frequencies from 250 to 8000 Hz in both ears. The CI group consisted of six post-lingually deafened adult CI users, who were tested with their clinical processors. The daily settings of volume control and microphone sensitivity were used throughout the tests. Table I lists the CI subject demographics. All CI subjects had more than one year of CI use and were tested with a single CI. Bilaterally implanted CI2 was asked to only use his first implant for testing. Bimodal CI users CI1, CI4, and CI6 were tested with the hearing aid removed and a foam earplug inserted in the non-implanted ear to prevent any contribution from residual acoustic hearing. CI3 and CI5 were unilateral CI users. Only two NH listeners (NH1 and NH3) and three CI users (CI1, CI2, and CI6) had spent a few years learning and playing a musical instrument before college. No subject

(either NH or CI) was engaged in musical activities in the past three years. All subjects provided informed consent and were paid for their participation. This study was approved by the Purdue IRB committee.

Before the experiments, CI users completed a familiar melody recognition test. The 12 familiar melodies used for melodic interval rating and adjustment (see below for details) were played with the original pitch and rhythm cues three times in random order. Subjects identified the melodies by clicking on response buttons marked with the titles of the melodies. The percent correct scores for familiar melody recognition are listed in Table I for reference.

B. Stimuli

The harmonic complex tones used in this study contained the first 15 consecutive harmonics (including the fundamental frequency; F0) in sine phase with a -12 dB/octave amplitude roll-off. The durations of complex tones were 400 ms in the melodic interval ranking test, but varied for familiar melodies based on the rhythm cues in the melodic interval rating and adjustment test. To better simulate real musical instrument sounds such as piano notes, the temporal envelopes of complex tones were modified to have an exponential decay with a time constant of 200 ms before being gated with 20-ms onset and offset raised-cosine ramps. All the complex tones were generated digitally at a 22 050-Hz sampling rate with 16-bit resolution in MATLAB and presented at 70 dBA via the basic loudspeaker of a GSI-61 audiometer in a double-walled, sound-treated booth.

Figure 1 shows the electrodiagrams for two pairs of complex tones or two melodic intervals that were representative of the stimuli used in the melodic interval ranking test (see below for details). The F0s of the two notes increased from 174.6 to 184.9 Hz for the first 1-semitone ascending interval and from 349.2 to 466.1 Hz for the second 5-semitone ascending interval. The notes were processed by the default ACE strategy of the Nucleus CI devices with 22 channels, 8 maxima, and 900 pulses per second per channel. It can be seen that the pulse trains on individual electrodes contained amplitude modulations at the F0s for the first but not the second interval. Thus, CI processors only provided temporal cues for interval perception at low F0s. Perceptually, CI users were also less sensitive to such temporal pitch cues at higher F0s (e.g., Zeng, 2002). On the other hand, there were more place cues to perceive the second larger interval than the first smaller interval. When the F0 increased by 5 semitones in the second interval, a different set of electrodes

TABLE I. CI subject demographics.^a

Subject	Age (yrs)	Gender	Etiology	Implant	Strategy	Duration of CI use (yrs)	Familiar melody recognition (%)
CI1	84	F	Sudden Hearing Loss	HiRes90K	HiRes-120	4	91.67
CI2	30	M	Otosclerosis	Nucleus 5	ACE	2	100.00
CI3	49	M	Sudden Hearing Loss	Clarion 1.0	CIS	17	83.33
CI4	70	M	Noise Exposure	Nucleus 5	ACE	3	100.00
CI5	44	F	Sudden Hearing Loss	HiRes90K	HiRes-120	8	91.67
CI6	24	M	Genetic	HiRes90K	HiRes-120	2	N/A

^aACE: Advanced Combination Encoder; CIS: Continuous Interleaved Sampling.

(e.g., EL13 instead of EL14) were stimulated and the intensities of stimulation on some electrodes (e.g., EL16 and EL18) were increased to elicit a higher pitch. Although not visible in Fig. 1, the logarithmic frequency maps used in current CI devices typically assign more channels to higher frequency ranges than to lower frequency ranges (Singh *et al.*, 2009) and may thus provide more place cues for interval perception at higher F0s than at lower F0s. That is, as the F0 increases, a trade-off may exist between the place and temporal pitch cues for interval perception with clinical CI processors.

Note that the electric stimuli delivered to CI users in this study were determined in part by the CI processors rather than directly controlled by a research interface. As stated in Sec. IV, the present results represent the perception of melodic intervals based on the information transmitted by the CI processors (including not only pitch but also loudness cues), as opposed to interval perception based on purely pitch cues.

C. Basic pitch ranking and melodic interval ranking

Subjects were first tested with basic pitch ranking and then with melodic interval ranking. To facilitate the comparison with previous studies (e.g., Gfeller *et al.*, 2002; Kang *et al.*, 2009), basic pitch ranking thresholds were measured using a two-down/one-up adaptive procedure (Levitt, 1971) that converged to a difference in F0 with 70.7% correct responses. In each trial, there were two complex tones (separated by 1000 ms), one of which (randomly selected) had the standard F0, while the other had a higher F0 + $\Delta F0$. Subjects were asked to identify which tone was higher in pitch; visual feedback was provided. The standard F0 was roved over a range of ± 1.58 semitones around either F₃ (174.6 Hz) or F₄ (349.2 Hz). This frequency roving prevented the unwanted perceptual adaptation to the standard tone during pitch ranking (McDermott *et al.*, 2010). The $\Delta F0$ was initially large enough (3.16 semitones) for all subjects to reliably rank the two tones in pitch. The $\Delta F0$ was reduced after two consecutive correct responses and increased after an incorrect response. The $\Delta F0$ was initially adjusted by a factor of 2 and then by a factor of $\sqrt{2}$ after the first four reversals. The

adaptive procedure finished with ten reversals and the pitch ranking threshold was the geometric mean of $\Delta F0$ (in semitones) across the last six reversals. The two standard F0s (around 174.6 or 349.2 Hz) were tested three times in random order and the pitch ranking thresholds for each standard F0 were averaged across all three runs.

The testing procedure for melodic interval ranking was similar to that for basic pitch ranking. In each trial, there were two intervals or two pairs of complex tones, one of which (randomly selected) had the standard interval I, while the other had a larger interval I + ΔI . Subjects were asked to identify the larger interval; visual feedback was provided. The two complex tones of each interval were played back to back without any inter-stimulus time interval and the two intervals were separated by 1000 ms. In the ascending interval ranking test, the first tone had a lower F0 than the second tone in each interval; in the descending interval ranking test, the first tone had a higher F0 than the second tone in each interval. To test interval ranking in different pitch ranges, the lower tone of the first interval (also the root note of each trial) was roved over a range of ± 1.58 semitones around either F₃ (174.6 Hz) or F₄ (349.2 Hz). The lower tone of the second interval was roved over a range of 5–10 semitones above the lower tone of the first interval. The latter frequency roving range was chosen to be higher than the expected interval ranking thresholds even with CIs, so that the higher tone of the second interval was mostly higher in pitch than the higher tone of the first interval, no matter which interval was larger. Subjects thus could not identify the larger interval by simply comparing either the lower or higher tones of the two intervals. Instead, they had to compare the sizes of pitch change in the two intervals. To test the effect of standard interval size on interval ranking, 0, 1, 2, 3, 5, 7, or 9 semitones (i.e., those common to Western music) were used as the standard interval size in different conditions. The 0-semitone standard interval size meant that in each trial, one interval always had no pitch change. To identify the larger interval, subjects essentially judged which interval had pitch change while the other did not. This condition thus provided a measure of basic pitch discrimination (but not pitch ranking). In the other conditions with non-zero

FIG. 1. Electrodegrams or stimulation pulse trains on individual electrodes for two pairs of complex tones or two melodic intervals that were representative of the stimuli used in the melodic interval ranking test. The first interval increased in F0 from 174.6 to 184.9 Hz (one semitone, ascending), while the second interval increased in F0 from 349.2 to 466.1 Hz (five semitones, ascending). The notes were processed by the default ACE strategy of the Nucleus CI devices with 22 channels, 8 maxima, and 900 pulses per second per channel. In the Nucleus CI devices, lower-numbered electrodes are more close to the basal end of the cochlea, corresponding to higher frequencies. Electrodes 1 to 11 were not activated for these stimuli.

standard intervals, both intervals in each trial had pitch changes and subjects judged which interval had a larger pitch disparity than the other. The interval difference ΔI was initially large enough (11.1 semitones) for all subjects to reliably rank the two intervals. The ΔI (in semitones) was adjusted based on subject responses using the same adaptive rule used for ΔF_0 in basic pitch ranking. The adaptive procedure finished with ten reversals and the interval ranking threshold was the geometric mean of ΔI (in semitones) across the last six reversals. Together, there were two root notes (around 174.6 or 349.2 Hz) $\times$ two pitch directions (ascending or descending) $\times$ seven standard interval sizes (0, 1, 2, 3, 5, 7, or 9 semitones), resulting in a total of 28 conditions. These conditions were tested three times in random order and the interval ranking thresholds for each condition were averaged across all three runs.

D. Melodic interval rating and adjustment

To simulate everyday music listening, melodic interval rating and adjustment was tested with 12 familiar melodies (i.e., “Happy Birthday,” “Here Comes the Bride,” “Jolly Good Fellow,” “London Bridge,” “Lullaby and Goodnight,” “Mary Had a Little Lamb,” “Old McDonald,” “Star Spangled Banner,” “Take Me Out To the Ball Game,” “This Old Man,” “Twinkle, Twinkle, Little Star,” and “Yankee Doodle”). These melodies were selected from previous CI studies (e.g., Gfeller *et al.*, 2002; Kong *et al.*, 2004) and were familiar to all subjects in this study. Only the initial phrases (i.e., the first 12–18 notes) of the melodies were used for testing, with the F_0 s ranging from A^b_3 (207.6 Hz) to C_5 (523.2 Hz). The intervals between immediately successive notes in these melodies ranged from -8 to 12 semitones (see Fig. 2 for the histogram of intervals). The F_0 s and intervals in these melodies overlapped with those in the melodic interval ranking test. Rhythm cues were preserved for all the familiar melodies.

The 12 melodies were tested twice with different initial values of the interval scaling factor in random order, resulting in a total of 24 test conditions. In the first trial of each condition, the melody was played with all its intervals (in semitones) scaled by a random factor. In one condition (start-compressed), the scaling factor was initially below 1

FIG. 2. Histogram of melodic intervals (in semitones) for all the tested familiar melodies.

(i.e., an integer multiple of 0.1 between 0 and 0.5) so that at least NH listeners could hear that the intervals were compressed. In the other condition (start-expanded), the scaling factor was initially above 1 (i.e., an integer multiple of 0.1 between 2 and 3) so that at least NH listeners could hear that the intervals were expanded. The two conditions ensured that both compressed and expanded intervals were tested for each melody and allowed us to check whether the interval adjustment results would be biased by the initial value of the scaling factor. Examples of a compression factor of 0.5 and an expansion factor of 2 can be found in Figs. 3(b) and 3(c), respectively. After listening to the melody, subjects were asked to rate how in tune the melody was, using a scale from 0 (completely out of tune) to 100 (completely in tune). Subjects were then asked to adjust the interval scaling factor to correct the melody. They could increase or decrease the scaling factor if the melodic intervals were perceived as compressed or expanded, respectively. The name of the melody was displayed in the testing interface to facilitate the rating and adjustment of the melody. Depending on the subject’s choice, the scaling factor may be increased or decreased in steps of 0.1 or 0.2. Subjects were encouraged to make larger adjustments in the first few trials to speed up the search for the scaling factor for the most in-tune melody

FIG. 3. Example melody “Old McDonald” with a scaling factor of 1, 0.5, 2, and -1 (panels a to d, respectively). The solid lines show the pitches of individual notes in semitones above A^b_3 (207.6 Hz). The numbers at the top of each panel show the directions (+/−) and sizes (in semitones) of pitch changes between successive notes.

FIG. 4. Group mean ascending (upward triangles) and descending (downward triangles) interval ranking thresholds as a function of standard interval size, as well as group mean basic pitch ranking thresholds (diamonds), with root note around 174.6 Hz (top panels) and 349.2 Hz (bottom panels) for NH listeners (left panels) and CI users (right panels). Symbols represent the mean, while error bars represent the standard deviation across subjects. For clarity of illustration, error bars are shown in only one direction for interval ranking thresholds. Schematics of example stimuli used in a single trial of each ranking test are shown in the legend.

(e.g., 1 for NH listeners), but they were encouraged to make smaller adjustments in the last few trials to fine tune the search results. The scaling factor had a minimum value of 0 (i.e., was never negative) so that the melodic intervals were scaled without reversing the directions of pitch changes (i.e., from ascending to descending and vice versa; see Fig. 3(d) for an example scaling factor of -1.0). The melody was played again with the adjusted intervals. The melodic interval rating and adjustment was repeated until the melody sounded the most in tune. Subjects were encouraged to explore a wide range of scaling factors during the procedure. The interval scaling factor and the corresponding melodic interval rating of each trial were recorded.

III. RESULTS

A. Basic pitch ranking and melodic interval ranking

Figure 4 shows the group mean thresholds in semitones for basic pitch ranking and melodic interval ranking for NH listeners (left panels) and CI users (right panels). To better illustrate the inter-subject variability of pitch and interval ranking with CI speech processors, Fig. 5 shows individual CI users' basic pitch ranking and melodic interval ranking thresholds with different root notes, pitch directions, and standard interval sizes (one subject per panel). Overall, the expected differences in pitch and interval ranking thresholds between CI and NH subjects or between the two root notes were not found, while both subject groups had higher interval ranking thresholds with increasing standard interval size and for descending intervals than for ascending intervals. Pitch ranking thresholds (diamonds) were analyzed by a mixed-design analysis of variance (ANOVA) with subject group (CI or NH) as the between-subject factor and root note (around 174.6 or 349.2 Hz) as the within-subject factor. The analysis showed that basic pitch ranking thresholds were slightly, but not significantly higher for CI users than for NH

listeners ($F_{1,10} = 3.63$, $p = 0.086$). The trend that basic pitch ranking thresholds increased with root note also did not reach significance ($F_{1,10} = 3.62$, $p = 0.086$). There was no significant interaction between subject group and root note ($F_{1,10} = 0.94$, $p = 0.36$).

Melodic interval ranking thresholds (upward and downward triangles) were first fit with a linear mixed model that included all the factors (categorical factors: subject group, root note, and pitch direction; continuous factor: standard interval size) and their interactions. It was found that all the interactions were not significant with p values close to 1. Based on the Akaike information criterion, the interactions were then removed from the model to most efficiently fit the interval ranking thresholds. The final model revealed borderline significant effects of subject group ($t_{10} = 2.24$, $p = 0.049$) and root note ($t_{321} = 2.05$, $p = 0.041$). Interval ranking thresholds were slightly higher for CI users than for NH listeners and with 349.2 Hz than with 174.6 Hz. Subjects had significantly higher ranking thresholds for descending intervals than for ascending intervals ($t_{321} = 4.84$, $p < 0.001$). Also, interval ranking thresholds significantly increased with standard interval size ($t_{321} = 11.68$, $p < 0.001$). On average, there was an increase of 0.22 semitones in the interval ranking thresholds when the standard interval size increased by 1 semitone.

When the standard interval size was 0 semitones, the interval ranking task was in fact to identify the interval with pitch change, which was similar to a pitch discrimination task. For each subject group (NH listeners or CI users), the thresholds obtained in this and the basic pitch ranking tasks, with both root notes (around 174.6 or 349.2 Hz), were compared using a two-way repeated measures (RM) ANOVA. It was expected that the interval ranking thresholds with the standard interval size of 0 semitones would be lower than the basic pitch ranking thresholds, because a discrimination task is typically easier than a ranking task and CI subjects

FIG. 5. Ascending (upward triangles) and descending (downward triangles) melodic interval ranking thresholds as a function of standard interval size, as well as basic pitch ranking thresholds (diamonds), for individual CI users (one subject per panel), with root note around 174.6 Hz (filled symbols) and 349.2 Hz (open symbols).

may use non-pitch cues to detect a non-zero melodic interval. While individual notes were 400 ms in both tasks, the time interval between the two compared notes was longer in the basic pitch ranking task (1000 ms) than in the melodic interval ranking task (0 ms), which may also make the basic pitch ranking task more difficult. However, the analyses showed that thresholds were similar with different pitch perception tasks ($F_{2,10}=3.80$, $p=0.059$ for NH listeners and $F_{2,10}=2.09$, $p=0.17$ for CI users) and root notes

($F_{1,5}=0.37$, $p=0.57$ for NH listeners and $F_{1,5}=0.40$, $p=0.55$ for CI users). The two factors did not significantly interact with each other ($F_{2,10}=1.00$, $p=0.40$ for NH listeners and $F_{2,10}=1.86$, $p=0.21$ for CI users).

B. Melodic interval rating and adjustment

Figure 6 shows the representative data from a NH subject NH3 and a CI subject CI6 for melodic interval rating

FIG. 6. Representative data from a NH subject NH3 (filled symbols) and a CI subject CI6 (open symbols) for melodic interval rating and adjustment with the melody “Jolly Good Fellow.” The interval scaling factor (top panel) and its corresponding melodic interval rating (bottom panel) are shown as a function of trial number. Data in the start-expanded condition are plotted using upward triangles connected by solid lines, while data in the start-compressed condition are plotted using downward triangles connected by dashed lines. The gray line in the top panel indicates the scaling factor of 1, while the gray line in the bottom panel indicates the highest rating of 100 for melodic intervals completely in tune.

and adjustment with the melody “Jolly Good Fellow.” In both the start-expanded and start-compressed conditions, NH listeners generally adjusted the interval scaling factor toward 1 with a few reversals and chose a final scaling factor around 1; exceptionally large final scaling factors were mostly found in NH4. As the scaling factor was adjusted, the melodic interval ratings of NH listeners gradually increased up to 100 (i.e., completely in tune). Some scaling factors were tested more than once during the adjustment of the same melody. The interval ratings for the same scaling factor were not necessarily the same in different trials because the interval ratings were relative to those in the preceding trials. On the other hand, CI users did not choose a consistent final scaling factor even for the same melody in the start-

expanded and start-compressed conditions. As implied by the example in Fig. 6, the final scaling factors of CI users were sometimes around 1, but sometimes even larger than 2 or smaller than 0.5. Although CI users also gradually increased the melodic interval ratings as they adjusted the scaling factor, their final interval ratings were mostly lower than 100. CI users often complained that some intervals in the melodies were more in tune than the others, no matter how they adjusted the scaling factor. Overall, the melodies were similarly out of tune with a wide range of scaling factors for CI users.

For each subject, the distribution of interval ratings for the same scaling factor was analyzed across all the melodies in both the start-expanded and start-compressed conditions. The median interval rating is plotted as a function of log scaling factor for individual NH listeners in Fig. 7 and CI users in Fig. 8. For both NH1 and NH3, the median interval ratings had a single sharp peak at the scaling factor of 1 and quickly reduced on both sides of the peak (Fig. 7). Although the effect of musical experience on interval ratings was not (and could not be) explicitly studied in the small number of non-musician listeners, NH1 and NH3 happened to be the only two NH subjects who had spent a few years learning and playing piano before college. For the other NH listeners, their interval ratings for scaling factors below 1 were similar to those of NH1 and NH3. However, their interval ratings reduced more slowly with fluctuations for scaling factors above 1 than for scaling factors below 1 along the log scale, suggesting that they were less sensitive to interval expansions than to interval compressions. For example, NH4 and NH5 had surprisingly high interval ratings even for scaling factors as large as 2.5. To examine whether such high ratings were due to misunderstanding of the task, NH4 was retested after additional practice with trial-to-trial instructions. However, similar results were found in the retest. It was noted that NH4 finished the tests with fewer trials (and possibly with less stringent criteria) than the other NH listeners. In Fig. 8, the interval rating functions of CI users except that of CI4 were relatively flat without clear peaks. The width of each subject’s interval rating function was calculated to indicate how quickly the subject reduced the interval ratings with increasing and decreasing scaling factors. The smallest and largest scaling factors with an interval rating equal to or larger than the 75th percentile of the subject’s entire interval ratings were found. The distance between the two scaling factors along the log scale was calculated as the width of the interval rating function, over which the subject gave the top 25% of his/her interval ratings. A t -test showed that the functions were significantly wider for CI users than for NH listeners ($t_{10} = 2.37$, $p = 0.04$), which suggests that CI users were less sensitive to interval expansions and compressions than NH listeners. For both subject groups, the interval ratings were not higher for integer scaling factors (e.g., 2 and 3) than for non-integer factors, suggesting that subjects did not prefer the integer-semitone intervals that occur most commonly in actual melodies. The musical experience of our non-musician subjects may not be enough to result in different ratings for intervals of integer and non-integer semitones.

FIG. 7. Melodic interval rating as a function of interval scaling factor for individual NH listeners. Circles represent the median values while error bars represent the 25th and 75th percentiles of interval ratings for the same scaling factor across all the melodies in both the start-expanded and start-compressed conditions.

Figure 9 shows the final scaling factors and final interval ratings for each subject with different melodies in both the start-expanded and start-compressed conditions (open and filled circles, respectively). Being an objective measure of perceptual accuracy, the final scaling factor was of more interest than the subjective measure of final interval rating. Overall, there was no clear difference between the final scaling factors of CI and NH subjects or in the two test conditions, due to the large variability across subjects and melodies. The small difference between CI and NH performance on the melodic interval adjustment task also seems to be due to unexpectedly poor performance by some of the NH listeners (e.g., NH2 and NH4), rather than exceptionally good performance by the CI users. The final scaling factors were analyzed by a mixed-design ANOVA with subject

group (CI or NH) as the between-subject factor and melody (e.g., “Happy Birthday,” “Here Comes the Bride,” etc.) and test condition (start-expanded or start-compressed) as the within-subject factors. It was found that the final scaling factors of CI users did not significantly differ from those of NH listeners ($F_{1,10} = 0.69$, $p = 0.43$). The final scaling factors were also not significantly different in the two test conditions ($F_{1,10} = 3.27$, $p = 0.10$), although some subjects (NH4, CI1, and CI5) chose larger final scaling factors in the start-expanded condition than in the start-compressed condition. The effect of melody on final scaling factors was significant ($F_{11,108} = 2.08$, $p = 0.028$). However, *post hoc t*-tests with Bonferroni correction showed no significant difference between the final scaling factors of any two melodies. Thus, there did not seem to be any special melody that should be

FIG. 8. Same as Fig. 7, but for individual CI users.

FIG. 9. Final scaling factors (top panels) and final interval ratings (bottom panels) for individual CI (left panels) and NH subjects (right panels) with different melodies in both the start-compressed (filled circles) and start-expanded conditions (open circles).

included or excluded in a clinical test. Clinicians may choose a set of melodies that are familiar to the patients to explore the variability in interval rating and adjustment. None of the interactions among factors was significant for the final scaling factors ($p > 0.25$ for all the interactions).

To quantify the accuracy of melodic interval adjustment, the difference between the final scaling factor and the target scaling factor (i.e., 1) was calculated in the log (base 2) scale. The squared errors of the final scaling factors were analyzed by a mixed-design ANOVA similar to that used for the final scaling factors. According to the analysis, the squared errors of the final scaling factors were not significantly different between CI and NH subjects ($F_{1,10} = 2.28$, $p = 0.16$), in the two test conditions ($F_{1,10} = 4.07$, $p = 0.07$), or across different melodies ($F_{11,108} = 1.47$, $p = 0.16$). There was no significant interaction among these factors ($p > 0.27$) for the squared errors of the final scaling factors.

The final interval ratings were 100 for all NH subjects except NH6. However, the final interval ratings were mostly lower than 100 for CI users (especially for CI2, CI3, and CI6). Non-parametric Kruskal-Wallis one-way ANOVA tests were used to compare the final interval ratings across subject groups, melodies, and test conditions. The analyses showed that the final interval ratings of CI users were significantly lower than those of NH listeners ($\chi^2_1 = 131.89$, $p < 0.001$). The final interval ratings were not significantly different for different melodies ($\chi^2_{11} = 1.46$, $p = 1.00$) or in different test conditions ($\chi^2_1 = 1.013$, $p = 0.31$).

C. Correlations between different measures of interval perception

Spearman rank order correlations were calculated between the melodic interval ranking thresholds and measures of melodic interval rating and adjustment. Figure 10(a) shows that

the average ranking threshold for ascending intervals across different standard interval sizes and root notes was significantly correlated with the width of the interval rating function (calculated at the 75th percentile of a subject's entire interval ratings

FIG. 10. Width of the interval rating function (top panel) and RMS error of the final scaling factors (bottom panel) as a function of average ranking threshold for ascending intervals. The open circles show the data for NH listeners, while the filled circles show the data for CI users.

along the log scale of the scaling factor) for CI users ($\rho = 0.94$, $p = 0.017$), but not for NH listeners ($\rho = 0.66$, $p = 0.18$). The two measures were ranked in the same order across all NH subjects except NH6, who had the widest interval rating function but the third lowest ranking threshold for ascending intervals among NH subjects. As shown in Fig. 10(b), there was a significant correlation between the average ranking threshold for ascending intervals and the root-mean-square (RMS) error of the final scaling factors (calculated across different melodies and test conditions) for CI users ($\rho = 0.89$, $p = 0.033$), but not for NH listeners ($\rho = 0.83$, $p = 0.06$). For each subject, the RMS error of the final scaling factors indicated the accuracy of melodic interval adjustment. The lack of correlation for NH listeners was because NH4 had the greatest error of the final scaling factors while her ranking threshold for ascending intervals was the fourth lowest among NH subjects.

On the other hand, the average ranking threshold for descending intervals across different standard interval sizes and root notes was not correlated with the width of the interval rating function for either CI users ($\rho = 0.60$, $p = 0.24$) or NH listeners ($\rho = 0.66$, $p = 0.18$). Also, the average ranking threshold for descending intervals was not correlated with the RMS error of the final scaling factors for either CI users ($\rho = 0.43$, $p = 0.42$) or NH listeners ($\rho = 0.83$, $p = 0.06$).

IV. DISCUSSION

A. Basic pitch ranking and melodic interval ranking

The basic pitch ranking thresholds for non-musician NH listeners in this study were below 1 semitone, while their melodic interval ranking thresholds were on average above 1 semitone (the smallest tonal unit in Western music). The results were consistent with those in previous studies obtained with similar adaptive procedures (for basic pitch ranking: Wier *et al.*, 1977; Spiegel and Watson, 1984; Michey *et al.*, 2006; for melodic interval ranking: Burns and Ward, 1978; Burns and Campbell, 1994; McDermott *et al.*, 2010). McDermott *et al.* (2010) proposed that non-musician NH listeners may use the overlearned musical pitch structures (such as familiar melodies and scales) to overcome the seemingly inadequate interval ranking thresholds (i.e., those above 1 semitone) for music perception. Zarate *et al.* (2012) argued that even for musicians, 1 semitone may be the intervallic limit to interval perception induced by musical training.

CI users in this study had an exceptional ability of basic pitch ranking that was not significantly different from that of NH listeners. As shown in Fig. 5, six CI users all exceeded the expectations in pitch ranking performance. Previous studies (e.g., Gfeller *et al.*, 2002; Nimmons *et al.*, 2008; Kang *et al.*, 2009) showed that CI users' pitch ranking thresholds, although highly variable across subjects, were significantly poorer than those of NH listeners. For example, Kang *et al.* (2009) reported a mean pitch ranking threshold of 3 semitones (ranging from 1 to 8 semitones) in 42 CI users, as compared to the average threshold of 1 semitone in NH listeners. In contrast, CI users in this study had an average pitch ranking threshold of 0.77 semitones (ranging from 0.23 to 2.56 semitones), which were at the lower end of previously reported thresholds for CI users. As in previous

studies (e.g., Gfeller *et al.*, 2002; Nimmons *et al.*, 2008; Kang *et al.*, 2009), a convenience sample of volunteer CI users rather than a selected group based on musical experience or performance was tested in this study. There were no notable differences in CI subject demographics between this and previous studies. The lower pitch ranking thresholds of CI users in this study were reliable because similar results were found in the melodic interval ranking test with the standard interval size of 0 semitones. Note that the lower thresholds of CI users in this study even corresponded to a higher percentage point (70%) than those in Kang *et al.* (2009; 50%). The standard F0s were similar in the two studies. The minimum interval tested by Kang *et al.* (2009) was 1 semitone, which may have prevented the measurement of pitch ranking thresholds below 1 semitone in NH listeners (e.g., McDermott *et al.*, 2010) or even some CI users (as in this study). Similar to Kang *et al.* (2009), this study did not incorporate amplitude roving in pitch ranking, so that the task was not too difficult for CI users. This, however, raised the possibility that there were unintended loudness cues correlated with F0. For example, the Nucleus 5 processor has directional microphones with a response that increases with frequency. Although the Advanced Bionics processors did not have directional microphones, they instead had pre-emphasis with a response that also increases with frequency. Thus, an increase in F0 may cause an increase in signal amplitude of CI output, and in turn, an increase in loudness. The trial-to-trial feedback in this study (but not in Kang *et al.*, 2009) may help CI users learn to respond to the unintended loudness cues.

In this study, the difference in melodic interval ranking thresholds between NH listeners and CI users was also not as big as expected, consistent with the similar basic pitch ranking performance of the two groups. For both NH listeners and CI users, the basic pitch ranking thresholds were similar to the melodic interval ranking thresholds with the standard interval size of 0 semitones. When the standard interval in the interval ranking test had no pitch change, the task to find the larger interval was simplified as a pitch discrimination task to find the interval with pitch change (with no need to identify the direction of pitch change). Although it may be easier to discriminate than rank pitches, pitch discrimination tested in the form of interval ranking with four notes in each trial may have similar perceptual demands as pitch ranking with only two notes in each trial, leading to similar thresholds in both tests. When the standard interval size was not zero, comparing the interval sizes in the two pairs of complex tones may have required more complicated memory and processing of pitch information, resulting in higher interval ranking thresholds than those with the standard interval size of 0 semitones. In addition, the suspected loudness cues for basic pitch ranking or discrimination were less likely to contribute to the ranking of two non-zero melodic intervals. For both subject groups, interval ranking thresholds significantly increased with standard interval size. Vos and Troost (1989) found that in Western music, the proportional occurrence of melodic intervals peaked at 2 semitones and exponentially reduced with larger interval sizes (e.g., Fig. 2). The less often heard larger intervals (e.g., 5–9 semitones) may have less robust internal representations and thus

higher ranking thresholds than the more often heard smaller intervals (e.g., 1–3 semitones). Another possible explanation is that Weber's Law in human perception may also apply to melodic interval ranking so that the just-noticeable difference between intervals is proportional to the size of the standard interval and higher ranking thresholds are thus needed for larger standard intervals.

The basic pitch ranking and melodic interval ranking thresholds for both CI users and NH listeners were not greatly affected by the two root notes. Such CI results differed from those with single-electrode stimuli, which showed that basic pitch discrimination (e.g., Zeng, 2002; Chatterjee and Peng, 2008; Luo *et al.*, 2010) and melodic interval perception (e.g., Pijl, 1997; McDermott and McKay, 1997) with temporal pitch cues alone worsened with higher pulse rates or amplitude modulation frequencies. Note that this study tested CI users with their clinical processors, which encode both place and temporal pitch cues. As the F0 increases, there may be a trade-off between the place and temporal pitch cues via clinical processors (see the electro-dograms in Fig. 1 for examples). That is, temporal pitch cues may be degraded while place pitch cues may be enhanced with higher F0s (Singh *et al.*, 2009; Luo *et al.*, 2014). The data of individual CI users in Fig. 5 showed variable effects of root note on pitch and interval ranking. CI2, CI4, and CI6 tended to perform better with 174.6 Hz than with 349.2 Hz, and thus may have been more sensitive to temporal rather than place pitch cues. In contrast, CI1 generally had lower ranking thresholds with 349.2 Hz than with 174.6 Hz, showing a greater sensitivity to place rather than temporal pitch cues. For CI3 and CI5, the trade-off between place and temporal pitch cues may have led to overall similar performance with the two root notes. In addition, the large range of frequency roving used in the pitch and interval ranking tests may have also reduced the effect of root note.

To our knowledge, ranking thresholds for descending intervals have not been previously measured or compared with those for ascending intervals. In this study, interval ranking thresholds were significantly higher for descending intervals than for ascending intervals. The different ranking results for ascending and descending intervals were not related to their proportional occurrence in Western music. Vos and Troost (1989) found that musical intervals of smaller sizes are predominantly descending, while those of larger sizes are predominantly ascending. However, such interaction between pitch direction and interval size was not observed for interval ranking in this study. Interval processing at central levels beyond the cochlea may be responsible for the effects of pitch direction on interval ranking, because similar effects were observed for both NH listeners and CI users with completely different cochlear processing. The planum temporale and intraparietal sulcus in the posterior auditory cortex have been shown to exhibit more activity with larger pitch change and thus may be involved in interval processing (Rinne *et al.*, 2007; Hyde *et al.*, 2008; Zarate *et al.*, 2010). Future imaging studies should test whether greater cortical activity in the auditory cortex and intraparietal sulcus may be recruited for ascending intervals than for descending intervals. Another possibility is that the different

ranking thresholds for ascending and descending intervals were an artifact of the testing procedure. In each trial, there were two pairs of notes; let us call the four notes A, B, C, and D. In addition to the two intervals AB and CD to be compared, there was a third interval BC that was not intended to be of interest and had a one-second (instead of none) inter-stimulus time interval. The interval BC was much larger in the descending interval ranking task than in the ascending interval ranking task (see the legend of Fig. 4 for examples). Also, in the descending interval ranking task, the interval BC was always in the opposite direction (i.e., ascending). The higher ranking thresholds for descending intervals may indicate that it is more difficult to compare the sizes of two intervals AB and CD if there is a larger distractor interval BC between them. A test of this hypothesis is to exchange the order of the two intervals in the descending interval ranking task and see if the ranking thresholds would be reduced by this change in the testing procedure. A better procedure would be to have a sequence of three notes A, B, and C, equally spaced in time, and to ask listeners which interval is larger (AB or BC). This procedure would be more relevant to the experience of listening to a melody.

B. Melodic interval rating and adjustment

The melodic interval rating and adjustment test in this study may be a useful tool to evaluate interval perception in not only NH listeners but also CI users. Unlike the identification and production of isolated intervals according to Western music labels (e.g., Pijl, 1997; McDermott and McKay, 1997), the rating and adjustment of intervals in the context of melodies only required familiarity with the melodies and were suitable for testing subjects with limited musical knowledge, as shown in this study. Compared to the frequently used familiar melody recognition test, the interval rating and adjustment test revealed more details about subjects' ability to perceive intervals. As shown in Table I, CI users in this study performed surprisingly well in the familiar melody recognition test, when rhythm cues were available. Their performance was similar to those of NH listeners and much better than those of CI users in previous studies of familiar melody recognition with rhythm cues (e.g., Gfeller *et al.*, 2002; Kong *et al.*, 2004). This may be because CI users in this study had an exceptional ability of basic pitch ranking (see Sec. III) and have been tested with familiar melody recognition several times for the other projects in our lab (leading to potential learning effects). In contrast, the final interval ratings of CI users were significantly lower than those of NH listeners, suggesting that the familiar melodies sounded out of tune after CI processing. Thus, CI users in this study may have relied on rhythm and pitch contour cues rather than pitch interval cues to achieve near-normal familiar melody recognition. Also note that both the RMS error of the final scaling factors and the width of the interval rating function did not have ceiling effects as in the familiar melody recognition scores and better reflected the individual differences in interval perception among CI users and/or NH listeners.

As indicated by the correlation results, CI users who had lower or better average ranking thresholds for ascending

intervals tended to choose more accurate final scaling factors for different melodies in the two test conditions, and their interval ratings reduced more quickly from the peak values. The correlations were significant even for the small number (6) of CI users tested in this study. In contrast, the average ranking threshold for descending intervals was not correlated with either the RMS error of the final scaling factors or the width of the interval rating function. These results suggest that CI users may have focused or relied on ascending intervals rather than descending intervals during the rating and adjustment of familiar melodies. Given that the familiar melodies tested in this study had an overall similar number of ascending and descending intervals (Fig. 2), ascending intervals with significantly lower ranking thresholds (Fig. 4) may have provided more salient cues than descending intervals for CI users to rate and adjust the familiar melodies.

Compared to the melodic interval ranking test, the melodic interval rating and adjustment test had some advantages for clinical and research use. Subjects showed a greater interest in listening to ecologically relevant familiar melodies than comparing isolated intervals. On average, the melodic interval rating and adjustment test took only an hour, while the melodic interval ranking test took eight hours. Most importantly, the outcomes of melodic interval rating and adjustment (i.e., the RMS error of the final scaling factors and the width of the interval rating function) were significantly correlated with the average ranking threshold for ascending intervals. With these properties, the melodic interval rating and adjustment test may provide a relatively quick but informative assessment of a CI user's interval ranking ability with different processing strategies.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in this study. Research was supported in part by NIH Grant R21-DC-011844.

- Attneave, F., and Olson, R. K. (1971). "Pitch as a medium: A new approach to psychophysical scaling," *Am. J. Psychol.* **84**, 147–166.
- Burns, E. M., and Campbell, S. L. (1994). "Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception," *J. Acoust. Soc. Am.* **96**, 2704–2719.
- Burns, E. M., and Ward, W. D. (1978). "Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of the melodic musical intervals," *J. Acoust. Soc. Am.* **63**, 456–468.
- Chatterjee, M., and Peng, S.-C. (2008). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," *Hear. Res.* **235**, 143–156.
- Cooper, W. B., Tobey, E., and Loizou, P. C. (2008). "Music perception by cochlear implant and normal-hearing listeners as measured by the Montreal battery for evaluation of amusia," *Ear Hear.* **29**, 618–626.
- Cuddy, L. L., and Cohen, A. J. (1976). "Recognition of transposed melodic sequences," *Q. J. Exp. Psychol.* **28**, 255–270.
- Dowling, W. J. (1978). "Scale and contour: Two components of a theory of memory for melodies," *Psychol. Rev.* **85**, 341–354.
- Dowling, W. J., and Fujitani, D. S. (1970). "Contour, interval, and pitch recognition in memory for melodies," *J. Acoust. Soc. Am.* **49**, 524–531.
- Galvin, J. J., Fu, Q.-J., and Nogaki, G. (2007). "Melodic contour identification by cochlear implant listeners," *Ear Hear.* **28**, 302–319.
- Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (2002). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," *Cochlear Implant Int.* **3**, 29–53.
- Hevner, K. (1935). "The affective character of the major and minor modes in music," *Am. J. Psychol.* **47**, 103–118.
- Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). "Evidence for the role of the right auditory cortex in fine pitch resolution," *Neuropsychologia* **46**, 632–639.
- Kang, R., Nimmons, G. L., Drennan, W., Longnion, L., Ruffin, C., Nie, K., Won, J. H., Worman, T., Yueh, B., and Rubinstein, J. (2009). "Development and validation of the University of Washington Clinical Assessment of Music Perception test," *Ear Hear.* **30**, 411–418.
- 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.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Looi, V., McDermott, H. J., McKay, C. M., and Hickson, L. (2008). "Music perception of cochlear implant users compared with that of hearing aid users," *Ear Hear.* **29**, 421–434.
- 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.
- Luo, X., Masterson, M. E., and Wu, C.-C. (2014). "Contour identification with pitch and loudness cues using cochlear implants," *J. Acoust. Soc. Am.* **135**, EL8–EL14.
- Maher, T. F. (1980). "A rigorous test of the proposition that musical intervals have different psychological effects," *Am. J. Psychol.* **93**, 309–327.
- McDermott, J. H., Keebler, M. V., Micheyl, C., and Oxenham, A. J. (2010). "Musical intervals and relative pitch: Frequency resolution, not interval resolution is special," *J. Acoust. Soc. Am.* **128**, 984–993.
- McDermott, H. J., and McKay, C. M. (1997). "Musical pitch perception with electrical stimulation of the cochlea," *J. Acoust. Soc. Am.* **101**, 1622–1631.
- Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). "Influence of musical and psychoacoustical training on pitch discrimination," *Hear. Res.* **219**, 36–47.
- Nimmons, G. L., Kang, R. S., Drennan, W. R., Longnion, J., Ruffin, C., Worman, T., Yueh, B., and Rubenstein, J. T. (2008). "Clinical assessment of music perception in cochlear implant listeners," *Otol. Neurotol.* **29**, 149–155.
- Peretz, I., Champod, A., and Hyde, K. (2003). "Varieties of musical disorders: The Montreal battery of evaluation of amusia," *Ann. N.Y. Acad. Sci.* **999**, 58–75.
- Pijl, S. (1997). "Labeling of musical interval size by cochlear implant patients and normally hearing subjects," *Ear Hear.* **18**, 364–372.
- Pijl, S., and Schwarz, D. W. (1995). "Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes," *J. Acoust. Soc. Am.* **98**, 886–895.
- Rakowski, A. (1990). "Intonation variants of musical intervals in isolation and in musical contexts," *Psychol. Music* **18**, 60–72.
- Rinne, T., Kirjavainen, S., Salonen, O., Degerman, A., Kang, X., Woods, D. L., and Alho, K. (2007). "Distributed cortical networks for focused auditory attention and distraction," *Neurosci. Lett.* **416**, 247–251.
- Russo, F. A., and Thompson, W. F. (2005). "The subjective size of melodic intervals over a two-octave range," *Psychon. Bull. Rev.* **12**, 1068–1075.
- Siegel, J. A., and Siegel, W. (1977). "Absolute identification of notes and intervals by musicians," *Percept. Psychophys.* **21**, 399–407.
- 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**, 160–168.
- Spiegel, M. F., and Watson, C. S. (1984). "Performance on frequency-discrimination tasks by musicians and non-musicians," *J. Acoust. Soc. Am.* **76**, 1690–1695.
- Swanson, B. A. (2008). "Pitch perception with cochlear implants," Ph.D. thesis, University of Melbourne, Melbourne, Australia.
- Vos, P., and Troost, J. (1989). "Ascending and descending melodic intervals: Statistical findings and their perceptual relevance," *Music Percept.* **6**, 383–396.
- Wang, W., Zhou, N., and Xu, L. (2011). "Musical pitch and lexical tone perception with cochlear implants," *Int. J. Aud.* **50**, 270–278.

- Ward, W. D. (1954). "Subjective musical pitch," J. Acoust. Soc. Am. **26**, 363–380.
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.
- Zarate, J., Ritson, C., and Poeppel, D. (2012). "Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary?," J. Acoust. Soc. Am. **132**, 984–993.
- Zarate, J. M., Wood, S., and Zatorre, R. J. (2010). "Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers," Neuropsychologia **48**, 607–618.
- Zatorre, R. J., and Halpern, A. R. (1979). "Identification, discrimination, and selective adaptation of simultaneous musical intervals," Percept. Psychophys. **26**, 384–395.
- Zeng, F.-G. (2002). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.