Influence of pitch, timbre and timing cues on melodic contour identification with a competing masker (L)

Meimei Zhu and Bing Chen[a)]
Department of Otorhinolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, 83 Fenyang Road, Shanghai 200031, People's Republic of China

John J. Galvin III and Qian-Jie Fu
Division of Communication and Auditory Neuroscience, House Ear Institute, Los Angeles, California 90057

(Received 19 April 2011; revised 11 October 2011; accepted 12 October 2011)

Pitch, timbre, and/or timing cues may be used to stream and segregate competing musical melodies and instruments. In this study, melodic contour identification was measured in cochlear implant (CI) and normal-hearing (NH) listeners, with and without a competing masker; timing, pitch, and timbre cues were varied between the masker and target contour. NH performance was near-perfect across different conditions. CI performance was significantly poorer than that of NH listeners. While some CI subjects were able to use or combine timing, pitch and/or timbre cues, most were not, reflecting poor segregation due to poor spectral resolution. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3658474]

PACS number(s): 43.66.Jh, 43.66.Mk, 43.66.Ts, 43.64.Me [DD] Pages: 3562–3565

I. INTRODUCTION

While cochlear implants (CIs) provide good speech understanding to many profoundly deaf individuals, challenging listening conditions (e.g., speech perception in noise, music perception) are difficult for CI users. Music perception is more difficult for CI users when music is played by multiple instruments (e.g., Gfeller *et al.*, 2002; Looi *et al.*, 2008; McDermott, 2004). Listeners must use pitch, timbre, and/or timing cues to segregate the melodic and rhythmic components ("analytic listening") and to stream groups of instruments ("synthetic listening"). Indeed, music perception and appreciation is enhanced by the shifting weights of analytic and synthetic listening: melodic lines emerge from dense chord structures; syncopated rhythms arise from multiple percussion lines, etc.

Rhythm, pitch, and timbre are three basic cues used for music perception. CI users' rhythm perception is nearly as good as that of normal hearing (NH) listeners (e.g., Kong *et al.*, 2004; Gfeller and Lansing, 1991; McDermott, 2004). However, CI users have much greater difficulty with pitch and timbre perception (e.g., Gfeller *et al.*, 2002; Kong *et al.*, 2004; Galvin *et al.*, 2007; Galvin *et al.*, 2008). Pitch, timbre, and timing cues allow listeners to stream and segregate different melodic lines, rhythms, and chord structures (Bregman, 1999). Timing cues are critical to streaming and segregation. Simultaneous auditory objects are likely to be grouped together while timing differences may enhance segregation of auditory objects. In both cases, pitch and timbre cues may be used to segregate competing auditory objects. Because of relatively poor pitch and timbre perception, CI users have difficulty in forming auditory streams. While some studies have shown that CI users can stream and segregate simple sequences using place and rate pitch cues (Chatterjee *et al.*, 2006; Hong and Turner, 2006, 2009), others have shown that CI users cannot segregate auditory streams (Cooper and Roberts, 2009).

Recently, Galvin *et al.* (2009a) measured melodic contour identification (MCI) with a competing masker in NH and CI listeners. Listeners were asked to identify a target 5-note melodic contour presented simultaneously with a 5-note masker. The masker consisted of the same five notes, played with the same or different pitch and/or timbre as the target contour. NH listeners were generally unaffected by the maskers, suggesting strong auditory stream segregation. CI users were generally unable to utilize pitch and/or timbre cues, although musically experienced subjects seemed best able to use these cues. CI performance with the masker was much poorer than without, reflecting difficulties CI listeners must experience when listening to multi-instrument music, rather than music performed by one instrument. In Galvin *et al.* (2009a), the masker and target were presented simultaneously. Given that multi-instrument music often involves nonsimultaneous notes, and given CI listeners' difficulties using pitch and timbre cues, nonsimultaneous presentation of the masker and target may allow for better MCI performance. In this study, MCI performance with a competing masker was evaluated in NH and CI subjects. As in Galvin *et al.* (2009a), pitch and timbre cues were varied between the target and masker. In addition, the timing between the masker and target was varied to be simultaneous (as in Galvin *et al.*, 2009a), overlapping, or sequential. We hypothesized that nonsimultaneous presentation would provide better segregation of the target contour, and that CI subjects may better utilize pitch and timbre cues with the overlapping and/or sequential timing.

II. METHODS

A. Subjects

14 CI users (8 male, 6 female) and 12 NH listeners (5 male, 7 female) participated in this study. Table I shows the CI subject demographics. CI subjects were required to be

[a)]Author to whom correspondence should be addressed. Electronic mail: eent_chen@yahoo.com

 0001-4966/2011/130(6)/3562/4/$30.00

TABLE I. CI subject demographics.

Subject	Gender	Age at HL onset (yr)	Age at testing (yr)	CI experience (yr)	L/R	Etiology	Device/strategy	Music experience
CI1	F	10	61	16(R) 3.5(L)	R, L	Drug	AB/F120(R, L)	Yes
CI2	M	24	59	4	R	Genetic	Freedom/ACE	Yes
CI3	M	67	73	1.3	R	Noise	Freedom/ACE	Yes
CI4	M	2	25	0.75	L	Meningitis	Nucleus 5/ACE	No
CI5	M	31	61	11	L	Unknown	AB/F120	No
CI6	F	5	67	7	R	Genetic	Freedom/ACE	No
CI7	F	57	63	5	L	Unknown	Freedom/ACE	No
CI8	F	43	78	10	R	Unknown	Freedom/ACE	No
CI9	F	23	26	3	R	Unknown	Freedom/ACE	No
CI10	M	60	80	15	L	Noise	Nucleus 22/SPEAK	No
CI11	M	2	23	1.5(R) 15(L)	R, L	Genetic	Nucleus 5/ACE(R), Nucleus 22/SPEAK(L)	No
CI12	F	10	23	5	L	Unknown	Freedom/ACE	No
CI13	M	0	23	20	R	Genetic	Nucleus 22/SPEAK	No
CI14	M	12	35	23	L	Meningitis	Nucleus 22/SPEAK	Yes

18 years of age or older, with warble-tone thresholds <25 dB for audiometric frequencies between 250 and 4 000 Hz while wearing their clinical speech processors and settings. To avoid floor performance effects, CI subjects were also required to score >30% correct in the no masker condition. The mean age of CI subjects was 49.7 yr (range: 23–80 yr). Two subjects participated in previous MCI experiments (Galvin *et al.*, 2009a), and another two were musically experienced. The remaining 10 CI subjects had no musical training or previous MCI experience. The mean age of NH subjects was 35.3 yr (range: 22–54 yr). NH subjects were required to be 18 years of age or older with pure tone thresholds <15 dB for audiometric frequencies between 250 and 4000 Hz. One of the NH subjects was a musician and another had some musical training (piano and violin) during childhood. The remaining 10 NH listeners had no musical experience. None of NH subjects had previous MCI experience. Both CI users and NH subjects were paid for their participation and all provided informed consent before testing was begun, in accordance with the local Internal Review Board.

B. Stimuli

Target stimuli consisted of 5-note melodic contours similar to those used in Galvin *et al.* (2007, 2008, 2009a,b): rising, rising-flat, rising-falling, flat-rising, flat, flat-falling, falling-flat, falling-flat, and falling. Target contours were generated in relation to a "root note" (the lowest note in the contour), according to $f_n = 2^{n/12} f_{\text{ref}}$, where f_n is the frequency of the target note, n is the number of semitones relative to the root note, and f_{ref} is the frequency of the root note. For all target contours, the root note was A3 (220 Hz) and the instrument was the piano. The note duration was 300 ms and the duration between notes was 300 ms. To see how performance was affected by the pitch range within the target contour, the spacing between successive notes was varied to be 1, 2, or 3 semitones. Thus, for the rising contour, the notes would be A3, A#3, B3, C4, and C#4 with 1-semitone spacing, A3, B3, C#4, D#4, and F4 with 2-semitone spacing, and A3, C4, D#4, F#4, and A4 with 3-semitone spacing. The target was always played by a piano, the most difficult instrument in a previous MCI experiment (Galvin *et al.*, 2008).

Masker conditions were varied to provide different timing, pitch, and timbre cues that could be used for contour segregation, as illustrated in Fig. 1. The masker consisted of five identical notes (i.e., a flat contour). As with the target, the masker note duration was 300 ms and the duration between notes was 300 ms. The *masker timing* was varied to be presented at the same time as the target (simultaneous), 150 ms before the target (overlapping), or 300 ms before the target (sequential). In the sequential condition, the masker and target notes alternated every 300 ms. The *masker pitch* was varied to have the same (A3) or a higher root note (A5) as the target. Given the maximum pitch range of the target with 3 semitone spacing, the A5 pitch would not overlap with the target at all. The *masker timbre* was varied to be the same (piano) or different instrument (organ) as the target. The piano masker shared the same spectral and temporal properties as the piano target (sharp attack, gradual decay, complex, and irregular harmonics), while the organ did not (smooth attack and decay, regular harmonic pattern), as illustrated in Galvin *et al.* (2009b). Both masker and target contours were played by sampled instruments with MIDI synthesis (Roland Sound Canvas with Microsoft Wavetable synthesis); the piano sample was "Piano 1" and the organ sample "Organ 1." The long-term RMS amplitude was the same for the masker and target contours (65 dB). In total, there were 13 test conditions: MCI without a masker, MCI with a masker (3 timing × 2 timbre × 2 pitch). For each condition, 54 stimuli were presented (9 contours × 3 semitone spacings × 2 repeats), and each condition was tested twice. Thus, each stimulus was presented 4 times for each condition.

C. Procedure

All stimuli were presented at 65 dBA in sound field via a single loudspeaker located directly in front of the listener (1 m away). CI subjects were tested while using their clinical processors and settings. Before testing, subjects were given a

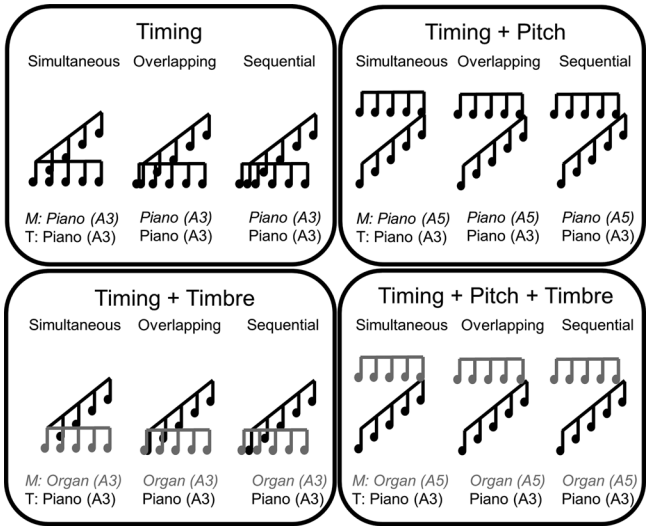

FIG. 1. Available cues for contour segregation. In the top left panel, only timing cues are available. In the top right panel, timing and pitch cues are available. In the bottom left panel, timing and timbre cues are available. In the bottom right panel, timing, pitch and timbre cues are available. In these examples, the target was rising (with 3 semitone spacing and A3 root note). The masker was flat with the same or different timing (simultaneous, overlapping, or sequential), pitch (A3 or A5) or timbre (piano or organ) as the target.

quick preview of the stimuli to familiarize them with the test procedures. Subjects were told that the masker would consist of the same note repeated 5 times, played by the piano or organ. During testing, a stimulus was randomly selected from the set and presented to the subject, who responded by clicking on one of nine response boxes labeled (with text and picture) according to the nine target contours. Subjects were allowed to repeat the stimulus a maximum of three times. No trial-by-trial feedback was provided. Test conditions were randomized within and across subjects. Each condition was tested twice and the mean performance for each subject was calculated.

III. RESULTS

Figure 2 shows mean MCI performance for CI users (left panel) and NH subjects (right panel); across subject means were calculated from two runs for each subject. Mean NH performance was near perfect across different condition with (96.0% correct) or without the masker (96.6% correct). When performance was collapsed across all conditions, a Kruskal– Wallis one way analysis of variance (ANOVA) on ranks showed that NH performance was significantly better than CI performance ($H = 220.893$; dF = 1; $p < 0.001$). When performance was collapsed across masker conditions, a Kruskal–Wallis one way ANOVA on ranks showed no significant difference between the no masker and masker conditions ($H = 0.853$; dF = 1; $p = 0.356$), most likely due to ceiling performance effects. Given the small decrement in performance with the sequential piano (A3) masker, we conducted a three-way repeated measures (RM) ANOVA. Results showed significant effects for timing [$F(2,121) = 8.56$, $p = 0.0003$], pitch [$F(1,121) = 13.70$, $p = 0.0003$], and timbre [$F(1,121) = 9.52$, $p = 0.0025$]. *Post hoc* Bonferroni pairwise comparisons showed that performance was significantly poorer only with the sequential timing when pitch and timbre cues were unavailable.

With no masker, mean CI performance was 61.0% correct (range: 37.0%–97.2% correct; standard deviation: 21.7%). With a masker, mean CI performance was 47.9% correct (range: 8.4%–97.2% correct; standard deviation: 25.6%). The effects of masker timing, pitch, and timbre were highly variable in CI users. The most masking occurred when the masker and target contours had the same timing (simultaneous), pitch (A3), and timbre (piano); mean performance decreased by 21.5 percentage points, relative to the no masker condition. The least masking occurred when the masker and target pitch were the same (A3), but the timbre (organ) and timing (sequential) were different; mean MCI performance decreased by 5.0 percentage points, relative to the no masker condition. When performance was collapsed across masker conditions, a Kruskal–Wallis one way ANOVA on ranks showed a significant difference in CI performance between the no masker and masker conditions ($H = 5.049$; dF = 1; $p = 0.025$). A three-way RM ANOVA showed no significant effect for timing, pitch, or timbre for CI users. However, there were significant interactions between timing and pitch [$F(2,143) = 4.85$, $p = 0.009$], and between pitch and timbre [$F(1,143) = 12.23$, $p = 0.0006$].

CI performance worsened as the semitone spacing in the target contours was reduced. A two-way RM ANOVA showed significant effects for semitone spacing [$F(2,312) = 9.608$, $p < 0.001$] and masker condition [$F(12,312) = 4.216$, $p < 0.001$]; there was no significant interaction. *Post hoc* Bonferroni pairwise comparisons showed that performance was significantly better with the 3-semtione spacing than with the 1-semitone spacing (adjusted $p < 0.001$).

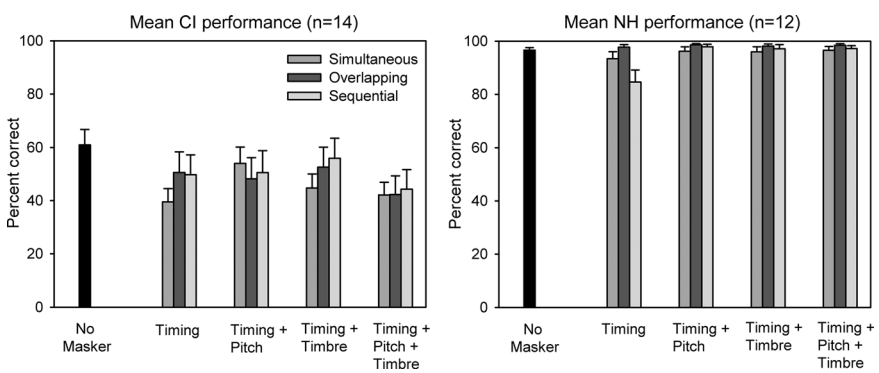

FIG. 2. Mean MCI performance for 14 CI subjects (left panel) and 12 NH subjects (right panel), as a function of masker condition. The black bars show performance with no masker; the remaining bars show performance for the different masker timing conditions. The error bars show one standard error.

Multiple linear regression analyses were performed to see whether CI subject demographics were correlated to MCI performance. For each masker condition, neither age at testing nor age at implantation predicted CI performance ($p > 0.05$). Linear regression analyses were also performed between the MCI data (with or without a masker) and speech data (HINT sentence recognition in noise, phoneme recognition in quiet) that were collected as part of a general CI speech performance database (12 of 14 CI subjects). MCI performance with the masker was significantly correlated with HINT speech reception thresholds (SRTs) and phoneme recognition (all p values < 0.05); MCI performance with no masker was significantly correlated only with HINT SRTs ($p < 0.05$).

IV. DISCUSSION

On average, CI subjects were unable to utilize or combine timing, pitch, and timbre cues to segregate the competing contours, similar to the results of Galvin *et al.* (2009a). While NH performance was nearly perfect with or without the masker, mean CI performance was significantly poorer with the masker than without. However, there was large inter-subject variability in the CI data, with some subjects scoring better than 95% correct in some masker conditions and others scoring less than 40% correct with no masker. Still, whether a good or poor performer, there was no consistent utilization of timing, pitch, and timbre cues among CI subjects.

The mean age of NH subjects was 35.3 yr (range: 22–54 yr), while the mean age of CI subjects was 49.7 yr (range: 23–80 yr). Age has been shown to be a factor in perception of spectrally degraded speech (as perceived by CI users), with poorer speech performance in older listeners (Schvartz *et al.*, 2008). Neither CI users' age at testing nor age at implantation was significantly correlated with MCI performance, with or without a masker. It is possible that a greater number of subjects might reveal a stronger influence of aging on MCI performance. Interestingly, CI users' sentence and phoneme recognition was significantly correlated with MCI performance with the masker. CI performance without the masker was correlated only with sentence recognition, different from the results of Galvin *et al.* (2007), which showed that MCI performance with a simple 3-tone complex was significantly correlated with vowel recognition.

Similar to previous CI speech perception experiments (e.g., Fu and Nogaki, 2004; Nelson *et al.*, 2003), CI subjects were unable to "listen in the dips" when the masker and target were nonsimultaneous. There was no significant difference between the overlapping and sequential timing conditions in this study, nor among the noise gating frequencies in Fu and Nogaki (2004). Indeed, this is the unfortunate new finding in the present study: even when provided with timing cues, CI listeners were unable to make use of pitch and/or timbre cues. For some listeners, the spectral resolution is so poor (and/or the channel interaction so great) that even when presented non-simultaneously, the masker and target cannot be segregated. Until the spectral resolution is greatly improved, CI subjects will most likely have great difficulty with multi-instrument music.

ACKNOWLEDGMENTS

We are grateful to all the subjects who participated to this experiment. This work was partially supported by NIH Grant No. R01-004993 and the National Natural Science Foundation of China (Grant No. 30872867).

Bregman, A. S. (**1999**). *Auditory Scene Analysis, the Perceptual Organization of Sound* (The MIT Press, Cambridge), Chap. 2, pp. 125–130.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (**2006**). "Auditory stream segregation with cochlear implants: A preliminary report," Hear. Res. **222**, 100–107.

Cooper, H. R., and Roberts, B. (**2009**). "Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition," J. Acoust. Soc. Am. **126**, 1975–1987.

Fu, Q.-J., and Nogaki, G. (**2004**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Galvin, J. J., 3rd, Fu, Q. J., and Nogaki, G. (**2007**). "Melodic contour identification by cochlear implant listeners," Ear. Hear. **28**, 302–319.

Galvin, J. J., 3rd, 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.

Galvin, J. J., 3rd, Fu, Q. J., and Oba, S. (**2009a**). "Effect of a competing instrument on melodic contour identification by cochlear implant users," J. Acoust. Soc. Am. **125**, EL98–EL103.

Galvin, J. J., 3rd, Fu, Q. J., and Shannon, R. V. (**2009b**). "Melodic contour identification and music perception by cochlear implant users," Ann. N.Y. Acad. Sci. **1169**, 518–533.

Gfeller, K., and Lansing, C. R. (**1991**). "Melodic, rhythmic, and timbral perception of adult cochlear implant users," J. Speech. Hear. Res. **34**, 916–920.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., and Knutson, J. (**2002**). "Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal," Ann. Otol. Rhinol. Laryngol. **111**, 349–356.

Hong, R. S., and Turner, C. W. (**2006**). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," J. Acoust. Soc. Am. **120**, 360–374.

Hong, R. S., and Turner, C. W. (**2009**). "Sequential stream segregation using temporal periodicity cues in cochlear implant recipients," J. Acoust. Soc. Am. **126**, 291–299.

Kong, Y. Y., Cruz, R., Jones, J. A., and Zheng, F. G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear. Hear. **25**, 173–185.

Looi, V., McDermott, H., McKay, C., and Hickson, L. (**2008**). "Music perception of cochlear implant users compared with that of hearing aid users," Ear. Hear. **29**, 421–434.

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

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc Am. **113**, 961–968.

Schvartz, K. C., Chatterjee, M., and Gordon-Salant, Sandra (**2008**). "Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners," J. Acoust. Soc. Am. **124**, 3972–3988.