Relative Contribution of Temporal and Spectral Cues to Melodic Contour Identification

Lichuan Ping[1, 2], Meng Yuan[1], Qinglin Meng[1, 2] and Haihong Feng[1]

1 Shanghai Acoustics Laboratory, Institute of Acoustics, Chinese Academy of Sciences, Shanghai, 200032, China
2 Graduate University of Chinese Academy of Sciences, Beijing, 100190, China

***Abstract*—The aim of this study was to investigate the relative contribution of temporal and spectral cues to melodic contour identification (MCI). Ten normal-hearing subjects participated in this study. Acoustic stimuli from four instruments and synthetic complex tone were utilized. A noise-excited vocoder was used to extract the temporal and spectral information from the stimuli. Experimental results showed that the spectral cue was more important to MCI than the temporal cue. Eight frequency bands were sufficient to achieve relatively good MCI performance. Temporal resolution showed no significant effect on the MCI performance. Confusion matrix was used to analysis the data. It was indicated that detecting pitch difference was easier than identifying pitch direction with restricted spectral information.**

Keywords- Cochlear implants; Temporal and Spectral Cues; Melodic Contour Identification

I. INTRODUCTION

Melody is one of the most important features of music. The perception of melody requires fine discrimination of changes in pitch, including both the direction and the degree of pitch change. Melodic contour is the relative up and down change in pitch. Galvin et al tested nine normal hearing (NH) and eleven cochlear implant (CI) subjects with melodic contour identification (MCI) task [1]. In his study, the subjects were asked to identify one of nine 5-note melodic contours. Results showed that the NH subjects achieved a mean score of about 95% correct, but the CI users' performance varied from 14% to 90% correct. He suggested that the poor MCI performance by CI subjects was probably caused by the limited spectral and temporal information delivered from the CI system.

In current CI devices, the signal processing strategies are all based on the "vocoder-centric" method [2-4]. In such method, the broad-band signal is filtered into several frequency bands. The slowly-varying temporal envelope of the band-limited signal is extracted from each band and used to modulate a fixed-rate pulse carrier. In these strategies, spectral resolution is limited by the number of available electrodes (8-22). The temporal resolution is limited by the cutoff frequency (typically 200 or 400 Hz) of the extracted envelope in each band.

Acoustic simulation of CI provides a method to flexibly control the temporal and spectral information of the stimuli [5-7]. Kong and her colleagues tested six NH subjects to investigate the contribution of spectral cues to melody recognition (without rhythmic) with the temporal envelope below 500 Hz [8]. The melody stimuli were processed to extract the envelope from 1 to 64 bands. The results showed that 32 bands were required for good melody recognition with NH subjects.

In the present study, psychoacoustic experiment was carried out using acoustic simulation of CIs to investigate the relative contribution of temporal and spectral cues to MCI performance. Stimuli from four instruments and synthetic complex tone were utilized. A noise-excited vocoder was used to control the spectral resolution (4, 8 and 16 frequency bands) and the temporal resolution (50 and 500Hz cutoff frequency of temporal envelope) of the stimuli. Ten NH subjects were tested with processed stimuli. Confusion matrix was used to analysis the data.

II. EXPERIMENTAL DESIGN

A. Subjects and Test Materials

Ten NH subjects participated in the experiment aging from 22 to 26. None of them had hearing disease before. Their audiometric thresholds were better than 20 dB HL at octave frequencies from 125 and 8000 Hz in both ears.

Figure 1 shows the nine melodic contours used in this study: Rising, Flat, Falling, Flat-Rising, Falling-Rising, Rising-Flat, Falling-Flat, Rising-Falling, and Flat-Falling.

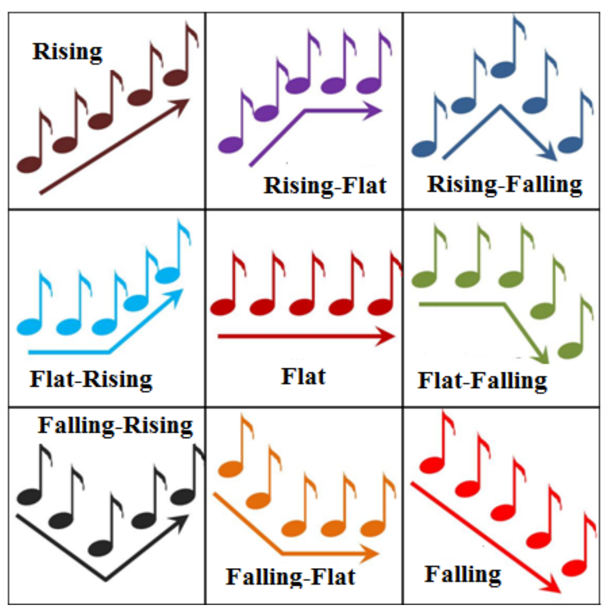

Figure 1. Nine melodic contours.

Each contour included five successive notes. The interval between adjacent notes varied between 1 and 6 semitones. The melodic contours were generated according to the first (lowest) note in the contour. The first note used in this study was C#4 (277Hz). There were four instruments, piano, trumpet, clarinet and violin, used in this study. They were selected because of the representation of different instrumental families. All musical tones were generated using the Cubase software (Steinberg Cubase VST Score 5.0). The synthetic complex tones were also included in this study. They were composed of F0 and two harmonics (that is, the 2nd and 3rd) similar to the "3-tone complex" in Galvin's MCI study [1]. The two harmonics had a successive 20% decrease in amplitude. The duration of all the stimuli was 500 ms with 25 ms on- and off-ramp, respectively. The intensity of all the stimuli was energy-balanced using RMS (Root Mean Square) equalization. The interval between adjacent notes was 500 ms.

B. Signal processing

A noise-excited vocoder was used to extract the temporal and spectral information of the stimuli [9-10]. The original stimuli were first band-pass filtered into several contiguous frequency channels using 6th-order elliptic filters. The frequency range was between 125 and 7999 Hz. The filters were designed using Greenwood function [11]. To avoid differences in group delay, zero-phase digital filtering was performed. The temporal envelopes were extracted by half-wave rectification and low-pass filtering (using 7th-order elliptic filters). The temporal envelopes were then used to amplitude-modulate independent white-noise carriers within the same bandwidth as the band-limited signal. Finally, acoustic stimuli were generated by combining the modulated signals from each band. The spectral resolution was controlled by the number of frequency bands (4, 8, and 16 bands) and the temporal resolution was controlled by the cutoff frequency of envelope (50 and 500Hz).

C. Test procedure

For each subject, six combinations of the temporal and spectral resolutions were tested in random order. The combinations were from two cutoff-frequencies (50 and 500 Hz) of the envelope and three frequency bands (4, 8, and 16 bands). Each combination contained five timbres (piano, trumpet, clarinet, violin and 3-tone complex) and two intonations (1 and 6 semitones interval). The five timbre conditions were tested in random order. In each timbre, there were 18 stimuli (9 contours * 2 intonations) which formed a test block. In each test block, a stimulus was randomly selected (without replacement) from the 18 stimuli and played to the subject. The subjects were asked to click on one of the nine response choices (grids in Figure 1) shown on the laptop screen. Each stimulus was allowed to repeat up to three times demanded by the subjects. The final score were reported as a percentage of correct responses.

The experiment was carried out in a sound-proof booth. The test equipment included a laptop computer with a high quality external audio interface (TASCAM US-122). The stimuli were played to the listeners via a Paired E.A.R. Tone 3A Insert Earphone (50Ω) at their comfortable levels. A graphical user interface was developed to control the presentation of the test stimuli and collect responses from subjects. A training session (15-20 minutes) was conducted for each subject first. Feedback was given to the subjects in this session.

III. RESULTS

The percentage correct of MCI performance was evaluated over all subjects. Figure 2 and 3 show the mean MCI performance of NH subjects with the 1- and 6-semitone intonation, respectively. The mean performance was shown for different combinations of spectral and temporal information. Different timbres were presented separately. The bold line in each box represents the median MCI performance. The chance level of the test was 11.11% correct.

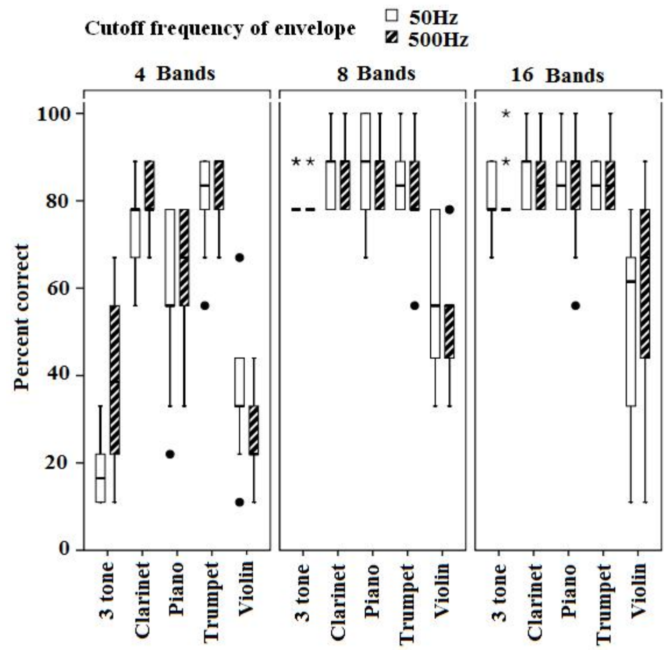

Figure 2. The MCI performance with the 1-semitone intonation.

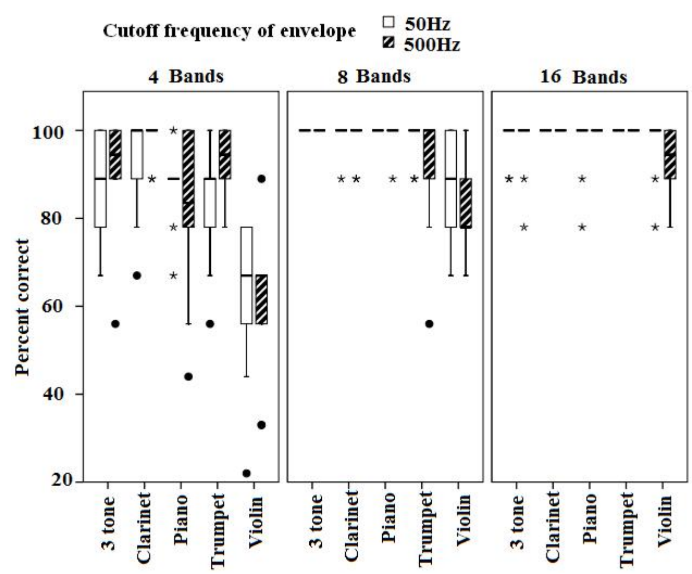

Figure 3. The MCI performance with the 6-semitone intonation.

In both figures, the scores varied from 11.11% to 100%. In general, the MCI performance was better with the increase of the number of frequency bands and cutoff frequency of envelope. The plateau performance for MCI was reached when the number of frequency bands was 8. The performance with the 6-semitone intonation was better than the 1-semitone one. Nearly perfect performance was achieved with 8 and 16 bands with the 6-semitone intonation. For different timbres, the MCI performance was the worst with violin.

The results were analyzed using three-way ANOVA, with the three factors: number of frequency bands (4, 8 and 16), cutoff frequency of envelope (50 and 500Hz), and timbre (piano, trumpet, clarinet, violin and 3-tone complex). The statistical analyses revealed significant main effects only for number of bands ([$F(4,10)=6.45$, $p=0.0078$]). No main effects were shown on the interactions among the three factors ($p>0.05$).

IV. DISCUSSION

In consistent with Galvin's MCI study [1], the performance was better with the increase of the semitone intonation. In a follow-up study, Galvin et al further investigated how the timbre cues interact with the CI users' melodic pitch perception [12]. The timbre cue was also observed to affect the MCI performance in that study, but not significant. In Kong's study, good melody recognition performance (>90) could be achieved with 32 frequency bands with the temporal envelope below 500 Hz [8]. In the current study, only 8 frequency bands were needed to achieve relatively good MCI performance (>80%) in most conditions. It may be due to the simple melodic structure compared to Kong's work. In this study, the melodic contour contains only 5 notes and the interval between consecutive notes was consistent.

It was shown that the MCI performance was the worst with 4 frequency bands in Figure 2 and 3. One needs to perceive the relative change in pitch direction for the identification of a melodic contour. Limited information provided by 4 frequency bands may not be adequate to support melodic contour identification.

Figure 4 shows the confusion matrix for the results with 4 frequency bands. Other factors, such as timbre, intonation and cutoff frequency of envelope were averaged.

The matrix was 9-by-9. The rows and columns present the user's response and the correct answers (the contour presented), respectively. The diagonal elements in the matrix represent the percentage correct MCI performance for each contour. It was shown that the Flat contour was recognized most accurately (93% correct), while the Rise and Fall contours were the worst (61% and 42% correct). The other six contours contain turning points where the melodic motion changes the direction. The results of these contours were all nearly 70% correct. The anti-diagonal of the matrix showed that the subjects tended to confuse the direction of melodic motion change. It's indicated that the subjects could detect the pitch difference more easily than identity the pitch direction with limited spectral information.

V. CONCLUSION

The results of the present study showed that the spectral cue was more important to MCI than the temporal cue. Eight frequency bands were sufficient to achieve relatively good MCI performance. Temporal resolution showed no significant effect on the MCI performance. The results indicated that detecting pitch difference was easier than identifying pitch direction with restricted spectral information.

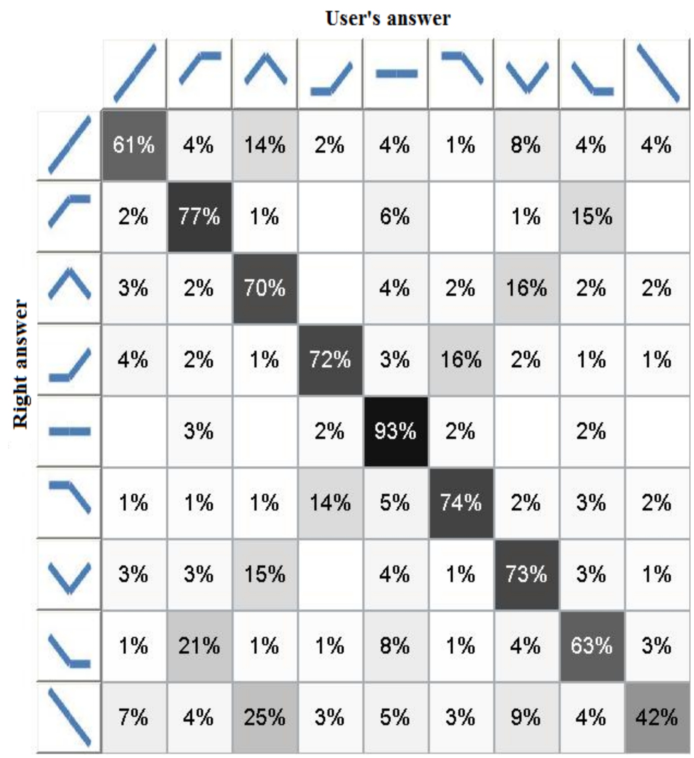

Figure 4. The confusion matrix for the MCI performance with four frequency bands.

ACKNOWLEDGEMENT

We are grateful to all the NH listeners for participating in this study. This work has been supported in part by the Chinese Academy of Sciences Pilot Project of the Knowledge Innovation Program (KGCX2-YX-607), Key Projects in the National Science & Technology Pillar Program in the Eleventh Five-year Plan Period (2008BAI50B08) and Shanghai Natural Science Foundation (11ZR1446000).

REFERENCES

[1] Galvin, J., Fu, Q.-J., and Nogaki, G., "Melodic contour identification by cochlear implant listeners," Ear Hear., vol.28, pp.302–319, 2007.

[2] Ward R. Drennan and Jay T. Rubinstein, "Music perception in cochlear implant users and its relationship with psychophysical capabilities", J Rehabil Res Dev. vol. 45(5), pp. 779–789, 2008.

[3] McDermott, H. J., "Music perception with cochlear implants: a review", Trends Amplif, vol. 8, pp. 49–82, 2004.

[4] Loizou, P., "Speech processing in vocoder-centric cochlear implants", Adv. Otorhinolaryngol. Basel, Karger, vol.64, pp.109–143, 2006.

[5] Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., "Speech recognition with primarily temporal cues", Science, vol.270, pp. 303–304, 1995.

[6] Xu, L., et al., “Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses,” J. Acoust. Soc. Am., vol.112, pp. 247–258, 2002.

[7] Kong, Y.-Y., and Zeng, F.-G., “Temporal and spectral cues in Mandarin tone recognition”, J. Acoust. Soc. Am., vol.120, pp. 2830–2840, 2006.

[8] Kong, Y.-Y., Cruz, R., Jones, J. A., et al., “Music perception with temporal cues in acoustic and electric hearing”, Ear Hear., vol. 25, pp. 173–185, 2004.

[9] Meng Yuan, et al., “Cantonese tone recognition with enhanced temporal periodicity cues” J. Acoust. Soc. Am., vol.126, pp. 327-337, 2009.

[10] Green, T., Faulkner, A., and Rosen, S., “Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved sampling cochlear implants,” J. Acoust. Soc. Am., vol.112, pp. 2155–2164, 2002.

[11] Greenwood, D.D., “A cochlear frequency-position function for several species–29 years later”, J. Acoust. Soc. Am., vol.87, pp. 2592–2605, 1990.

[12] Galvin, J., Fu, Q.-J., and Oba, S., “Effect of instrument timbre on melodic contour identification by cochlear implant users,” J. Acoust. Soc. Am., vol.124, EL189–EL195, 2008.