Available online at www.sciencedirect.com

SCIENCE DIRECT®

Hearing Research 197 (2004) 87–95

HEARING RESEARCH

www.elsevier.com/locate/heares

Mandarin tone recognition in cochlear-implant subjects ☆

Chao-Gang Wei [a], Keli Cao [a], Fan-Gang Zeng [a,b,c,d,*]

[a] *Department of Otolaryngology, Head and Neck Surgery, Peking Union Medical College Hospital, Beijing 100730, China*
[b] *Department of Anatomy and Neurobiology, Head and Neck Surgery, University of California, Irvine, CA 92697, USA*
[c] *Department of Biomedical Engineering, Head and Neck Surgery, University of California, Irvine, CA 92697, USA*
[d] *Department of Cognitive Sciences and Otolaryngology, Head and Neck Surgery, University of California, Irvine, CA 92697, USA*

Received 13 April 2004; accepted 14 June 2004
Available online 25 August 2004

Abstract

This study examined tone recognition in five cochlear-implant subjects who were native speakers of Mandarin and used a Nucleus-22 device. Psychophysical experiments were conduced to measure rate discrimination in individual electrodes from the most apical to the most basal electrodes. The rate range was from 100 to 200 Hz, which corresponded to the range of variation in fundamental frequency for the tonal tokens used in this study. Speech recognition experiments were also conducted to measure tone recognition as function of the number of electrodes from a 1-electrode map to a 20-electrode map. Large individual variability was observed for both rate discrimination and tone recognition result: Average rate discrimination ranged between 0.2 and 1.2 (Weber's fraction) whereas tone recognition ranged between 30% and 70% correct. A highly significant correlation was found between rate discrimination and tone recognition with the 20-electrode map, but a non-significant correlation was observed with the 1-electrode map due to a floor effect in tone recognition. The present result supports the hypothesis that both spectral and temporal cues contribute to tone recognition. In addition, the present result shows that current cochlear-implant subjects produced significantly lower performance than acoustic simulations in normal-hearing subjects, suggesting that neither temporal nor spectral cues have been adequately and appropriately extracted and encoded in current cochlear implants. New designs are discussed to improve tone recognition in cochlear implant subjects.

Keywords: Cochlear implant; Tone recognition; Speech recognition; Rate discrimination; Electrode; Channel

1. Introduction

More than 60,000 hearing-impaired persons worldwide have received cochlear implants and derived benefits in daily life. Due to economical reasons, the overwhelming majority of current cochlear implant users live in western countries where the cochlear implant was developed and manufactured. Recently, there has been a surge in cochlear implantation in developing countries, particularly in China where several hundred persons received a single-electrode device (Zeng, 1995) and more than 1600 persons have received modern multi-electrode cochlear implants (Cao, 2004; Han, 2004). Because essentially all Chinese dialects, including Cantonese and Mandarin, are tonal languages, in which pitch variations convey lexically different meanings, it would be interesting to find out whether speech processing strategies in cochlear implants are adequate to encode this important tonal information. After all, at least one quarter of the world population speaks tonal language (e.g., Chinese, Thai, and Vietnamese).

☆ A preliminary version of the present paper was published in Chinese text, but none of the present tables and figures was published previously (Wei, C., Cao, K., Wang, Z., Zeng, F.G. 1999. Rate discrimination and tone recognition in mandarin-speaking cochlear-implant listeners. Chin J. Otorhinolaryngol. 34, 84–88.).

[*] Corresponding author. Tel.: +949-824-1539; fax: +949-824-5907. *E-mail address:* fzeng@uci.edu (F.-G. Zeng).

0378-5955/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.heares.2004.06.002

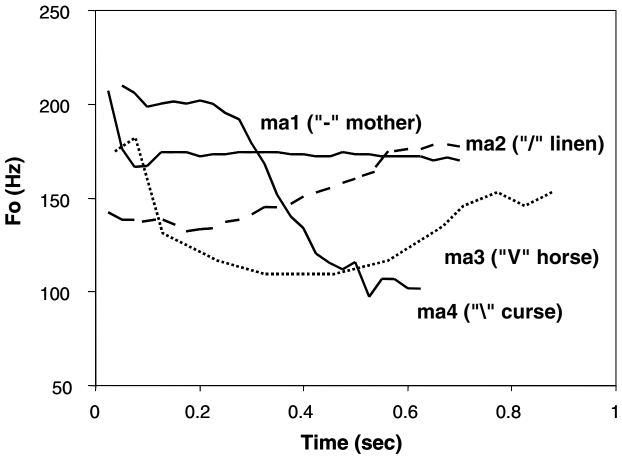

Fig. 1. Mandarin tone patterns. Fundamental frequency contours are plotted as a function of time. Tone 1 has flat and relatively high pitch contour ("-"), tone 2 has a rising contour, tone 3 has falling-rising contour, and tone 4 has a falling pattern. The longer duration for tone 3 and shorter duration for tone 4 are also typical for Mandarin tones.

Fig. 1 provides an example of the pitch variation patterns in Mandarin tones. The same consonant–vowel syllable, /ma/, may mean "mother", "linen", "horse", or "curse", depending on fundamental frequency variations including flat (tone 1), rising (tone 2), falling-rising (tone 3), and falling (tone 4), respectively. Although tonal patterns are typically illustrated by changes in fundamental frequency, acoustic analysis has shown that perceptual cues for tone recognition are widely distributed in both spectral and temporal domains.

Liang Zhi'an performed the first systematic study on acoustic cues in Mandarin tone recognition (Liang, 1963). He found that perfect tone recognition (90% correct or above) with a male talker could be achieved by either directly preserving the fundamental frequency with low-pass filtering at 300 Hz or indirectly by residual pitch derived from the harmonic structure with high-pass filtering at 300, 1200, and 2400 Hz. He further demonstrated that robust tone recognition (64% correct) could also be achieved by whispered speech, which presumably preserved only temporal and spectral envelope cues but did not contain any spectral fine structure related to fundamental frequency and its harmonics. Several follow-up studies further isolated the spectral envelope cues and found that the temporal envelope cues such as vowel duration and amplitude contours also contributed to Mandarin tone recognition (Blicher et al., 1990; Garding et al., 1986; Howie, 1974; Whalen and Xu, 1992). This contribution, while significant, was relatively weak when the more salient pitch cue evoked by fundamental frequency and its harmonics was present (Lin, 1988).

Two additional lines of research rekindled the discussion on relative contributions of spectral and temporal cues to tone recognition. One line of research was the recognition of the importance of temporal envelope cues in speech perception in general (Fu, 2002; Rosen, 1992; Shannon et al., 1995; Van Tasell et al., 1987). Several studies had since extended this line of work to Mandarin speech and found that the temporal envelope cue contributed significantly to both Mandarin tone and sentence recognition in normal-hearing listeners (Fu and Zeng, 2000; Fu et al., 1998; Kong et al., 2003; Xu and Pfingst, 2003). However, a recent study found that although the temporal envelope cue played a significant role in quiet listening condition, its practical role was significantly diminished in the presence of background noise (Kong et al., 2003).

Development of cochlear implants also stimulated research in studying acoustic and perceptual cues for tone recognition. Clinical studies have reported that Cantonese- and Mandarin-speaking cochlear-implant users had a great deal of difficulty in tone recognition, even after several years of device use when other aspects of speech perception had been significantly improved (Au, 2003; Huang et al., 1995; Lee et al., 2002; Wei et al., 1999, 2000; Wu and Yang, 2003). Generally, tone recognition reported in these studies ranged from 50% to 70% correct. The relatively poor performance in tone recognition had been linked to a number of factors including: (1) limitations in current cochlear-implant speech processing strategies that do not explicitly extract and encode the salient fundamental frequency and its harmonic cues (Lan et al., 2004; Nie et al., 2004; Zeng, 1995), (2) inadequate and even conflicting representations of the temporal envelope cues (Au, 2003; Green et al., 2002), and (3) limitation of temporal pitch to several hundred Hertz in electric hearing (Wei et al., 1999; Zeng, 2002).

One problem with previous studies on tone recognition was that they were either investigative but not direct (e.g., acoustic hearing experiments in normal-hearing subjects) or direct but mostly descriptive (e.g., electric hearing experiments in cochlear-implant subjects). The present study aimed to bridge this gap by conducting two experiments in rate discrimination and tone recognition as a function of number of active electrodes. The aim was to delineate relative contributions of temporal and spectral cues to tone recognition in a group of Mandarin-speaking cochlear-implant subjects.

2. Materials and methods

2.1. Subjects

Five cochlear-implant subjects who were native Mandarin speakers participated in the present study. These subjects received the Nucleus-22 device at the Cochlear Implant Center, Peking Union Medical College Hospital. Table 1 lists the subject information. Their age at

Table 1
Subject information

Subject	Age	Gender	Deaf dur. (years)	Etiology	CI use (mos)	Sentence (% correct)
C1	59	M	21	Ototoxity	30	61
C2	34	M	11	Ototoxity	30	97
C3	38	F	6	unknown	10	38
C4	44	F	24	unknown	10	4
C5	21	F	20	Ototoxity	22	0

the time of test ranged from 21 to 59 years. The cause of the deafness was ototoxity in three cases and unknown in the other two. While ototoxity-induced deafness is relatively uncommon in Western countries, ototoxity is still the most common cause for deafness in China (Zeng, 1995). Except for one (C5), all subjects were post-lingually deafened. Although three subject had extensive experience with an earlier version of the Nucleus MSP processor using the MPEAK strategy, they were using the spectra 22 processor with the SPEAK strategy at the time of test (Skinner et al., 1994). Open-set sentence recognition ranged from 0% (C5) to almost perfect (C2). The Local IRB approval and informed consent were obtained.

2.2. Stimuli

A 500-ms biphasic pulse train with 200-μs/phase was used in the rate discrimination experiment. Because the fundamental frequency varied between 100 and 200 Hz for the present tonal stimuli, two standard rates, 100 and 200 Hz, were used for the rate discrimination experiment. Test rates ranged from 103 to 392 Hz for the 100-Hz standard rate and from 203 to 496 Hz for the 200-Hz standard rate. The exact range was determined on an individual basis based on the implant user's ability to discriminate rate differences. The pulse train was delivered to a single electrode pair via the standard clinical interface. Four electrode pairs (BP+1 mode) were used, including from the most basal pair E1–3, E7–9, E14–16, to the most apical pair E20–22. All stimuli were presented at the most comfortable loudness level.

One-hundred tokens produced by a male talker were used in the Mandarin tone recognition experiment. These 100 tokens included 25 consonant–vowel combinations with each having four tonal patterns. The choice of these tokens were based on the following two principles, namely, maximizing the diversity of the consonant and vowel combinations and making sure that each consonant–vowel-tone combination is lexically meaningful in modern Chinese. Table 2 shows the complete set of stimuli used in the present study. The leftmost column shows the consonant and vowel combinations, whereas the rightmost four columns show the four tonal variations and their corresponding Chinese characters. A CD containing these 100 tokens along with a graphic user interface for testing is available by writing to the corresponding author.

Table 2
Mandorin tone stimuli

Phonemes	Tones			
	1	2	3	4
ma	妈	麻	马	骂
po	坡	婆	笸	迫
ke	科	壳	可	课
xi	西	习	洗	细
du	都	读	赌	度
ju	居	局	举	聚
cai	猜	才	彩	菜
wei	微	为	委	卫
bao	包	薄	饱	暴
chou	抽	愁	丑	臭
jia	家	夹	假	稼
jie	接	节	姐	借
qiao	敲	侨	巧	翘
you	优	由	有	又
duo	多	夺	躲	堕
hui	挥	回	恢	会
wan	弯	完	晚	万
shen	深	神	婶	甚
chang	昌	长	场	唱
feng	风	逢	讽	奉
tong	通	同	捅	痛
nian	粘	年	撵	念
pin	拼	贫	品	聘
liu	溜	流	柳	六
xuan	宣	悬	选	眩

2.3. Procedure

A method of constant stimuli was used to derive the difference limen in the rate discrimination experiment.

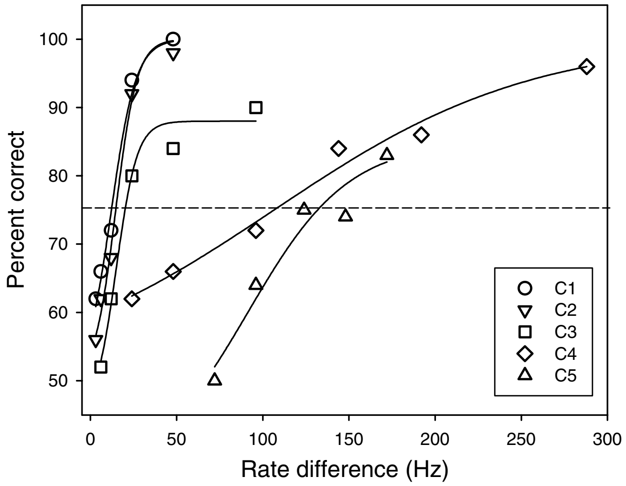

Fig. 2. Representative psychometric functions for rate discrimination. Percent correct is plotted as a function of rate difference with the 100-Hz standard rate in five individual cochlear-implant subjects (symbols). The solid lines represent best fitted sigmoidal functions to the individual data. The dashed line represents the criterion (75% correct) that determines the rate difference limen.

Typically, 5–6 test rates were used in combination with the standard rate to produce performance that ranged from chance (50%) to nearly perfect in most cases (occasionally the performance reached only 80% with the highest test rate). In each trial, a pair of stimuli was presented to the subject with one being the test rate and the other being the standard rate stimulus. Both the presentation order and sequence were randomized between trials. The subject's task was to tell whether one of the them had higher pitch. Each test rate and standard rate pair was presented 50 times and the percent correct was calculated. A psychometric function was generated with percent correct scores being plotted as a function of the difference between the test rate and the standard rate. A sigmoidal function was then fitted to the data to derive the difference limen that produced 75% correct response, corresponding to a d' value of 1. Weber's fraction, defined as the ratio between the difference limen and the standard rate, was reported in the study.

A four-alternative, forced-choice method was used in the tone recognition experiment. In each trial the subject was presented with a tone token in a random order and had to select the tone pattern (1, 2, 3, or 4) corresponding to the presented stimulus. The graphic interface presented both the phoneme-tone combinations and their corresponding Chinese characters. Five repetitions, representing a total of 500 tokens, were used to generate the percentage correct data reported in this study.

A total of eight experimental speech processor maps was generated, including four 1-electrode maps and four additional maps for 4, 7, 14, and 20 electrodes. The four 1-electrode maps were E1–3, E7–9, E14–16, and E20–22, corresponding to the same bipolar electrode pairs on which rate discrimination was measured. The additional four multiple-electrode maps were identical to the configurations used in a previous study (Fishman et al., 1997). For example, the 4-electrode map used E1–3, E7–9, E14–16, and E20–22 electrode pairs. Before the testing, each subject was given 30 minutes of practice to familiarize with the experimental processor. It should be noted that this short-term practice would not be enough to induce potential learning or training effect with these experimental processors (Rosen et al., 1999).

3. Results

3.1. Rate discrimination

Fig. 2 shows representative rate discrimination data at the 100-Hz standard rate in 5 cochlear-implant subjects. A sigmoidal function (solid line) was fitted to the data to derive the rate difference limen that was required to reach 75% correct response (dashed line). Note the great individual variability in the data that spans more than an order of magnitude: some subjects (e.g., C1) required only a 10-Hz difference whereas others (e.g., C5) required a 130-Hz difference to distinguish from the 100-Hz standard rate.

Table 3 presents individual Weber's fraction data collected from four electrode pairs at both 100- and 200-Hz standard rates. ANOVA was performed to find that neither the electrode position [$F(3,32) = 0.73$, $p > 0.5$] nor the standard rate [$F(1,32) = 1.11$, $p > 0.3$] was a significant factor. Therefore, the grand average over both electrodes and rates are calculated and later used for correlation with the tone recognition data. Fig. 3 shows such grand average data for rate discrimination in 5

Table 3
Weber's fraction for rate discrimination in five Mandarin-speaking Nucleus-22 cochlear-implant subjects

Subject	Standard rate = 100 Hz				Standard rate = 200 Hz			
	E1	E7	E14	E20	E1	E7	E14	E20
C1	0.50	1.22	0.12	0.22	0.29	0.36	0.67	0.66
C2	0.11	0.11	0.17	0.30	0.36	0.14	0.50	0.36
C3	0.14	0.30	0.17	0.37	0.05	0.10	0.11	0.35
C4	0.27	0.86	1.08	2.59	0.74	0.22	0.53	1.24
C5	1.75	1.40	1.50	1.25	0.83	0.63	1.23	1.18

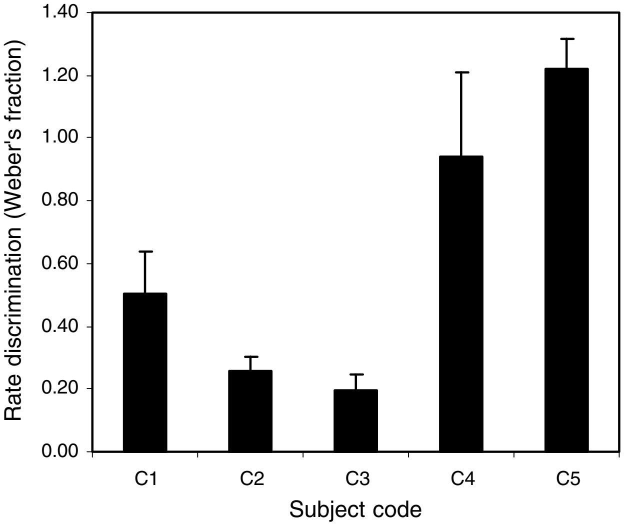

Fig. 3. Individual rate difference limens. The data are averaged across four bipolar electrode pairs and two standard rates. The error bars represent one standard error.

cochlear-implant subjects. Weber's fraction ranged from 0.20 to 1.22, with subject C3 having the best rate discriminability and subject C5 having the worst rate discriminability.

3.2. Tone recognition

Table 4 presents individual tone recognition data on four single-electrode maps (columns 2–5) and the averaged single-electrode result (column 6). Data points were missing on 3 of 4 single-electrode maps for subject C2 because the subject's complaint on the perceived sound quality on those electrodes, whereas no single-electrode data were collected for subject C5 because the subject's performance with the 20-electrode map was the chance level. With a few exceptions (e.g., E14 for subject C1), the single-electrode map produced essentially chance performance (25%).

Fig. 4 shows both the individual (symbols) and the mean (solid line) tone recognition data as a function of the number of electrodes. The single-electrode data were averaged across four single-electrode maps. For comparison, acoustic simulation data from normal-hearing listeners using the same stimuli and procedure are also presented as a function of the number of frequency bands (Kong et al., 2003). Several interesting observations can be made about the obtained results. Note first the monotonically increasing performance from 1-electrode to 7-electrode maps and the plateau in performance from 7- to 20-electrode maps. ANOVA confirmed the highly significant effect of the number of electrodes on tone recognition [$F(5,18) = 15.3$, $p < 0.001$] and the non-significant difference in performance between the 7- and 20-electrode maps (Scheffe post-hoc tests, $p > 0.5$). Note second the high level of performance in acoustic simulation of the cochlear implant, particularly with the low number of electrodes/bands. The difference between simulation and actual implant data was 45 percentage points with the 1-electrode condition and reduced to about 20 percentage points with high numbers of electrodes/bands. To amplify this point, we note that none of the actual implant subjects performed anywhere close to the simulation result with low numbers of electrodes/bands where the best implant performance approached the simulation performance with high numbers of electrodes/bands.

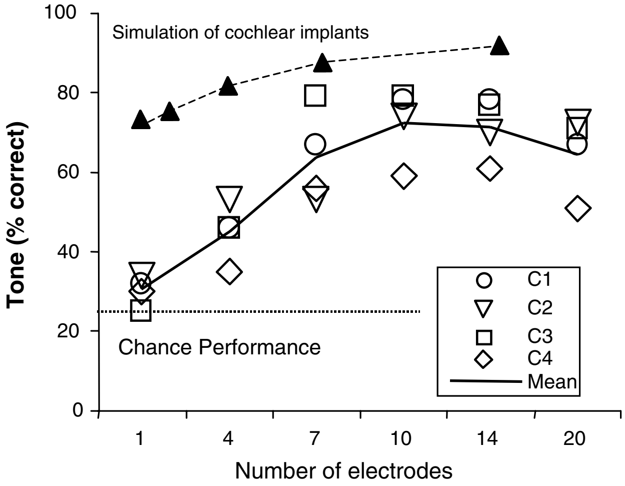

Fig. 4. Mandarin tone recognition as a function of the number of electrodes. Individual data are represented by symbols whereas the mean data are represented by the thick solid line. Chance performance is represented by the dotted line (25% correct). For comparison, acoustic simulation results as a function of the number of spectral bands are also included as triangles connected by a dashed line.

Table 4
Tone recognition with single-electrode experimental maps in five Mandarin-speaking Nucleus-22 cochlear-implant subjects. DNT denotes Did Not Test (see text)

Subject	E1	E7	E14	E20	Average
C1	26.0	35.0	42.0	24.0	31.8
C2	DNT	34.0	DNT	DNT	34.0
C3	31.0	24.0	18.0	27.0	25.0
C4	32.0	26.0	35.0	25.0	29.5
C5	DNT	DNT	DNT	DNT	N/A

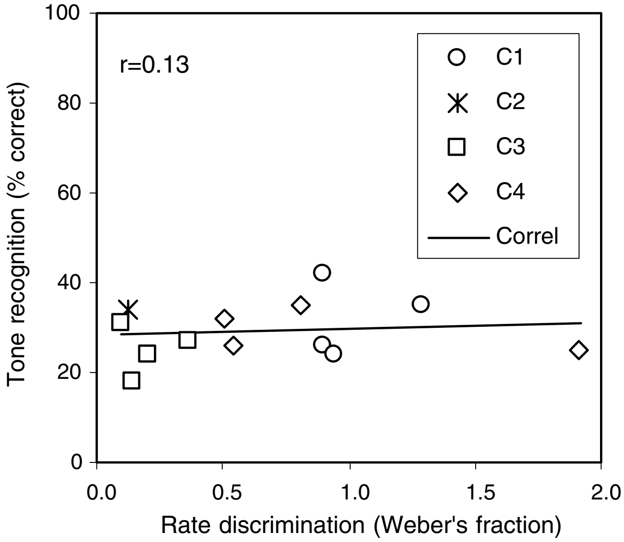

Fig. 5. Correlation between rate discrimination and tone recognition with 1-electrode maps. Individual data are presented by symbols whereas the regression function is represented by the solid line.

3.3. Correlational analysis

Fig. 5 shows the correlation between tone recognition and rate discrimination obtained with single electrodes. Although the rate discrimination varied greatly between different electrodes and among individual cochlear-implant subjects, all these single-electrode maps produced result close to chance performance, resulting in non-significant correlation between tone recognition and rate discrimination ($r = 0.13, p > 0.5$). On the contrary, a highly significant correlation ($r = -0.97$, $p < 0.01$) was observed between tone recognition with the full 20-electrode map and rate discrimination averaged over all 4 electrodes and 2 standard rates (top panel of Fig. 6). This significant correlation provided indirect evidence for the temporal contribution to tone recognition in cochlear-implant users. Additionally, due to the small number of subjects, only a trend in correlation ($r = -0.79$, $p = 0.06$) was observed between sentence recognition and rate discrimination (bottom panel of Fig. 6).

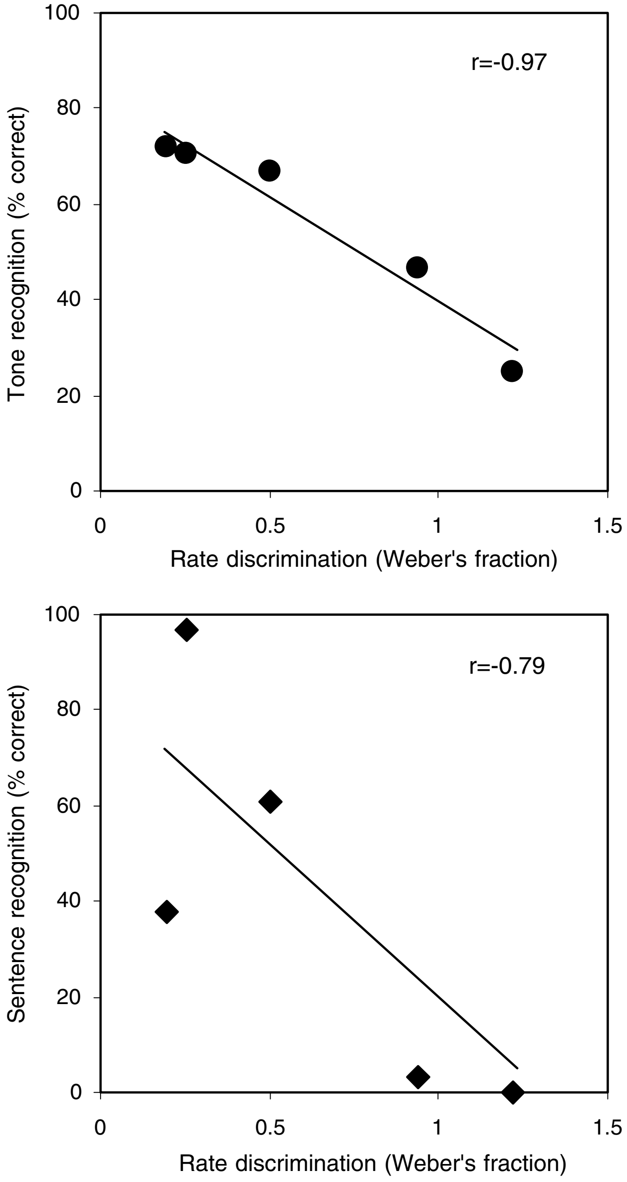

Fig. 6. Correlation between rate discrimination and tone recognition (top panel) and sentence recognition (bottom panel) with the 20-electrode map. Individual data are presented by symbols whereas the regression function is represented by the solid line.

4. Discussion

4.1. Comparison with previous data

The present rate discrimination data were in general agreement with previously reported data, for a summary, see (Zeng, 2002). The exception was the large Weber's fraction observed in the pre-lingually deafened subject, C5, suggesting that even rate pitch encoding may require significant auditory experience in early development. Similar suggestion has been made previously by showing that pre-lingually deafened cochlear-implant subjects could reliably estimate loudness but not pitch (Busby and Clark, 2000).

The present tone recognition data with the full 20-electrode map were also in agreement with previously reported data in cochlear-implant subjects who spoke tonal languages. For both Cantonese and Mandarin tone recognition, the previously reported data ranged from 50% to 70% correct (Au, 2003; Cao et al., 2000; Huang et al., 1995; Lee et al., 2002; Wei et al., 1999, 2000; Wu and Yang, 2003). The present 20-electrode map produced tone recognition with an average of

57% and a range from 25% to 71% correct performance. Additionally, the pattern of tone recognition was similar to previously reported English or German speech perception as a function of the number of electrodes: a monotonically increasing performance from 1 to 7 electrodes and a plateau in performance from 7 to 20 electrodes (Fishman et al., 1997; Friesen et al., 2001; Garnham et al., 2002).

However, there is a significant discrepancy between cochlear-implant performance as a function of electrode numbers and acoustic simulation performance as a function of spectral bands, particularly at low number of electrodes/bands (Fig. 4). When only the temporal envelope cues were available in the 1-band condition, previous acoustic simulations produced tone recognition performance ranging from 60% to 80% correct, depending on the cutoff frequency of the envelope extraction (Fu and Zeng, 2000; Fu et al., 1998; Kong et al., 2003; Xu et al., 2002). In contrast, the present cochlear-implant subjects only produced an average of 30% correct, barely above chance performance (25%). The causes for this discrepancy will be discussed in the following sections.

4.2. Spectral and temporal contributions to tone recognition

Previous studies have demonstrated that acoustic cues for tone recognition are widely distributed in both frequency and time domains (Liang, 1963; Lin, 1988; Whalen and Xu, 1992), and additionally the spectral and temporal cues can be traded for each other depending on relative availability and importance of these cues (Fu and Zeng, 2000; Xu and Pfingst, 2003). The present psychophysical and speech data in actual cochlear-implant subjects also demonstrated this interdependence and availability of acoustic cues for tone recognition. First, the low performance with the 1-electrode map suggests that current cochlear-implant subjects may likely have access to only durational cues. In acoustic simulations, the durational cues produced identical 30% correct performance; had the cochlear-implant subjects had access to additional amplitude and/or pitch periodicity cues, they would have obtained 60–70% correct performance (Fu and Zeng, 2000). Second, the increasing performance as a function of the number of electrodes suggests that spectral cues or temporal cues distributed in different spectral bands are important for tone recognition. It is possible that amplitude contour and pitch periodicity cues were somehow smeared under the 1-electrode condition when the temporal envelopes from different spectral bands were summed into 1 band. This view is supported by the present finding that rate discrimination was not correlated with the low performance in tone recognition from the 1-electrode map but was significantly correlated with the relatively high performance from the 20-electrode map. The plateau in performance between 7 and 20 electrode could be due to the conflicting cues between the periodicity cue coding tonal information and the amplitude cue coding formant information (Green et al., 2002). This hypothesis of conflicting cues may be supported by the trend that a peak performance was achieved with 10 electrodes coupled with a 10-percentage-point drop with 20 electrodes. Interestingly, similar trends were also observed for voicing identification in previous studies. For example, in both English and German consonant recognition, voicing identification peaked with 6–10 electrodes but dropped by 15 percentage points with 20 electrodes (Fishman et al., 1997; Garnham et al., 2002).

4.3. Tone encoding in cochlear implants

The most important finding in the present study is that the speech processing strategies in current cochlear implants do not provide sufficient acoustic cues to support adequate tone recognition. This result is also consistent with the finding that current cochlear implant subjects have a great deal of difficulty in recognizing music melodies, speakers, speech in noise, or other tasks requiring fine resolution of the pitch cue (Kong et al., 2004; Stickney et al., 2004; Vongphoe and Zeng, 2004). Several lines of research have been initiated to improve encoding pitch in cochlear implants.

The first line of research has been focused on the traditional temporal envelope domain. Attempts and suggestions have been made to improve or accentuate the temporal envelope representation of tonal information (Geurts and Wouters, 1999, 2001; Green et al., 2002; Litvak et al., 2003; Rubinstein and Hong, 2003). While these attempts should be able to do better than current devices, they are unlikely to solve ultimately the problem related to pitch coding of complex sounds because complex pitch perception requires correct tonotopic representation (Oxenham et al., 2004; Zeng, 2002).

Another line of research attempts to improve place coding of complex pitch perception by either increasing the spectral resolution at low frequencies (Geurts and Wouters, 2004) or aligning frequency-to-electrode maps to reflect the true place pitch in cochlear implants (Whitford et al., 1993). This line of inquiry is clearly needed but unfortunately no studies have been published yet to see whether improved tone recognition can be achieved with better place coding of fundamental frequency and its harmonics.

Finally, recent attempts have been made to improve pitch coding in temporal fine structure. One way to achieve this goal is to directly extract and encode fundamental frequency as a carrier for electric stimulation (Lan et al., 2004). The other way is to extract and encode frequency modulation in combination with amplitude modulation for electric stimulation (Nie et al., 2004). Similarly, the direct evidence for this line of research is lacking, but preliminary studies are promising. In a case report, a bilingual cochlear implant subject achieved significantly better performance with Chinese than with English (63% vs. 42% correct word recognition) with a now obsolete processing strategy (WSP) that directly encoded fundamental frequency (Xu et al., 1987). Acoustic simulation results encoding fundamental frequency and frequency modulations have also demonstrated significant improvement over the current speech processing strategies (Lan et al., 2004; Nie et al., 2004).

5. Summary

This study examined rate discrimination and tone recognition in five cochlear implant subjects who were native speakers of Mandarin and used a Nucleus-22 device. Weber's fraction for rate discrimination ranged between 0.2 and 1.2 whereas tone recognition ranged between 30% and 70% correct, corresponding to a 1-electrode map and a 20-electrode map, respectively. A highly significant negative correlation was found between rate discrimination and tone recognition with the 20-electrode map, but a non-significant correlation was observed with the 1-electrode map due to a floor effect in tone recognition. The present result supports the hypothesis that both spectral and temporal cues contribute to tone recognition, and additionally shows that neither temporal nor spectral cues have been adequately and appropriately extracted and encoded in current cochlear implants. New designs are needed to extract and encode explicitly fundamental frequency and its harmonics in order to improve tonal language recognition in cochlear-implant subjects.

Acknowledgement

The authors thank Zhizhong Wang for his support, Ting Zhang for preparing Fig. 1, and Lin Chen, Ying-Yee Kong and two anonymous reviewers for comments on the manuscript. This work was partially supported by National Natural Science Foundation of China (39570756 and 30228021) and National Institutes of Health (2RO1DC02267).

References

Au, D.K.K., 2003. Effects of stimulation rates on Cantonese lexical tone perception by cochlear implant users in Hong Kong. Clin. Otolaryngol. 28, 533–538.

Blicher, D.L., Diehl, R.L., Cohen, L.B., 1990. Effects of syllable duration on the recognition of the Mandarin Tone 2/Tone 3 distinction: evidence for auditory enhancement. J. Phonetics 18, 37–49.

Busby, P.A., Clark, G.M., 2000. Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant. J. Acoust. Soc. Am. 107, 547–558.

Cao, K., 2004. Standardization of cochlear implants. Chin. J. Otorhinolaryngol. 39, 71–72.

Cao, K., Wei, C., Wang, Z., 2000. Speech recognition and language development in prelingually deafened children with multichannel cochlear prosthesis. Zhonghua Er Bi Yan Hou Ke Za Zhi 35, 16–19.

Fishman, K.E., Shannon, R.V., Slattery, W.H., 1997. Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. J. Speech Lang. Hear. Res. 40, 1201–1215.

Friesen, L.M., Shannon, R.V., Baskent, D., Wang, X., 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am. 110, 1150–1163.

Fu, Q.J., 2002. Temporal processing and speech recognition in cochlear implant users. Neuroreport 13, 1635–1639.

Fu, Q.J., Zeng, F.G., 2000. Identification of temporal envelope cues in Chinese tone recognition. Asia Pacific J. Speech Lang. Hear. 5, 45–57.

Fu, Q.J., Zeng, F.G., Shannon, R.V., Soli, S.D., 1998. Importance of tonal envelope cues in Chinese speech recognition. J. Acoust. Soc. Am. 104, 505–510.

Garding, E., Kratochvil, P., Svantesson, J.O., Zhang, J., 1986. Tone 4 and Tone 3 discrimination in modern Standard Chinese. Lang. Speech 29, 281–293.

Garnham, C., O'Driscoll, M., Ramsden, R., Saeed, S., 2002. Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets. Ear. Hearing 23, 540–552.

Geurts, L., Wouters, J., 1999. Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants. J. Acoust. Soc. Am. 105, 2476–2484.

Geurts, L., Wouters, J., 2001. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J. Acoust. Soc. Am. 109, 713–726.

Geurts, L., Wouters, J., 2004. Better place-coding of the fundamental frequency in cochlear implants. J. Acoust. Soc. Am. 115, 844–852.

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

Han, D.M., 2004. Current status of multi-electrode cochlear implants in China. Chin. J. Otorhinolaryngol. 39, 70–71.

Howie, J.M., 1974. On the domain of tone in Mandarin. Phonetica 30, 129–148.

Huang, T.S., Wang, N.M., Liu, S.Y., 1995. Tone perception of Mandarin-speaking postlingually deaf implantees using the Nucleus 22-Channel Cochlear Mini System. Ann. Otol. Rhinol. Laryngol. Suppl. 166, 294–298.

Kong, Y.Y., Vongphoe, M., Zeng, F.G., 2003. Independent contributions of amplitude modulation and frequency modulation to auditory perception: II. Melody, tone and speaker identification. Abstract of the 26th annual midwinter research meeting 26, pp. 213–214.

Kong, Y.Y., Cruz, R., Jones, J.A., Zeng, F.G., 2004. Music perception with temporal cues in acoustic and electric hearing. Ear. Hearing 25, 173–185.

Lan, N., Nie, K.B., Gao, S.K., Zeng, F.G., 2004. A novel speech-processing strategy incorporating tonal information for cochlear implants. IEEE Trans. Biomed. Eng. 51, 752–760.

Lee, K.Y., van Hasselt, C.A., Chiu, S.N., Cheung, D.M., 2002. Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. Int. J. Pediatr. Otorhinolaryngol. 63, 137–147.

Liang, Z.A., 1963. Auditory perceptual cues in Mandarin tones. Acta Physiol. Sin. 26, 85–91.

Lin, M.C., 1988. Acoustic and perceptual cues for Mandarin tones. Chin. Linguist. 3, 182–193.

Litvak, L.M., Delgutte, B., Eddington, D.K., 2003. Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains. J. Acoust. Soc. Am. 114, 2079–2098.

Nie, K., Stickney, G.S., Zeng, F.G., 2004. Encoding fine structure to improve cochlear implant performance in noise. IEEE Trans. Biomed. Eng. (in press).

Oxenham, A.J., Bernstein, J.G., Penagos, H., 2004. Correct tonotopic representation is necessary for complex pitch perception. Proc. Natl. Acad. Sci. USA 101, 1421–1425.

Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B: Biol. Sci. 336, 367–373.

Rosen, S., Faulkner, A., Wilkinson, L., 1999. Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. J. Acoust. Soc. Am. 106, 3629–3636.

Rubinstein, J.T., Hong, R., 2003. Signal coding in cochlear implants: exploiting stochastic effects of electrical stimulation. Ann. Otol. Rhinol. Laryngol. Suppl. 191, 14–19.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.

Skinner, M.W., Clark, G.M., Whitford, L.A., Seligman, P.M., Staller, S.J., Shipp, D.B., Shallop, J.K., Everingham, C., Menapace, C.M., Arndt, P.L., et al., 1994. Evaluation of a new spectral peak coding strategy for the Nucleus 22 Channel Cochlear Implant System. Am. J. Otol. 15 (Suppl. 2), 15–27.

Stickney, G.S., Zeng, F.G., Litovsky, R.Y., Assmann, P.F., 2004. Cochlear implant speech recognition with speech masker. J. Acoust. Soc. Am. (in press).

Van Tasell, D.J., Soli, S.D., Kirby, V.M., Widin, G.P., 1987. Speech waveform envelope cues for consonant recognition. J. Acoust. Soc. Am. 82, 1152–1161.

Vongphoe, M., Zeng, F.G., 2004. Speaker recognition with temporal cues in acoustic and electric hearing. J. Acoust. Soc. Am. in press.

Wei, C., Cao, K., Wang, Z., Zeng, F.G., 1999. Rate discrimination and tone recognition in mandarin-speaking cochlear-implant listeners. Chin. J. Otorhinolaryngol. 34, 84–88.

Wei, W.I., Wong, R., Hui, Y., Au, D.K., Wong, B.Y., Ho, W.K., Tsang, A., Kung, P., Chung, E., 2000. Chinese tonal language rehabilitation following cochlear implantation in children. Acta Otolaryngol. 120, 218–221.

Whalen, D.H., Xu, Y., 1992. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica 49, 25–47.

Whitford, L.A., Seligman, P.M., Blamey, P.J., McDermott, H.J., Patrick, J.F., 1993. Comparison of current speech coding strategies. Adv. Otorhinolaryngol. 48, 85–90.

Wu, J.L., Yang, H.M., 2003. Speech perception of Mandarin Chinese speaking young children after cochlear implant use: effect of age at implantation. Int. J. Pediatr. Otorhinolaryngol. 67, 247–253.

Xu, L., Pfingst, B.E., 2003. Relative importance of temporal envelope and fine structure in lexical-tone perception. J. Acoust. Soc. Am. 114, 3024–3027.

Xu, L., Tsai, Y., Pfingst, B.E., 2002. Features of stimulation affecting tonal-speech perception: implications for cochlear prostheses. J. Acoust. Soc. Am. 112, 247–258.

Xu, S.A., Dowell, R.C., Clark, G.M., 1987. Results for Chinese and English in a multichannel cochlear implant patient. Ann. Otol. Rhinol. Laryngol. 96 (suppl 128), 126–127.

Zeng, F.G., 1995. Cochlear implants in China. Audiology 34, 61–75.

Zeng, F.G., 2002. Temporal pitch in electric hearing. Hearing Res. 174, 101–106.