Published in final edited form as:

J Acoust Soc Am. 2002 July ; 112(1): 247–258.

Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses

Li Xu^{a)}

Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109 and School of Hearing, Speech and Language Sciences, Ohio University, Grover Center, Athens, Ohio 45701

Yuhjung Tsai and Bryan E. Pfingst

Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, 1301 East Ann Street, Ann Arbor, Michigan 48109

Abstract

Tone languages differ from English in that the pitch pattern of a single-syllable word conveys lexical meaning. In the present study, dependence of tonal-speech perception on features of the stimulation was examined using an acoustic simulation of a CIS-type speech-processing strategy for cochlear prostheses. Contributions of spectral features of the speech signals were assessed by varying the number of filter bands, while contributions of temporal envelope features were assessed by varying the low-pass cutoff frequency used for extracting the amplitude envelopes. Ten normal-hearing native Mandarin Chinese speakers were tested. When the low-pass cutoff frequency was fixed at 512 Hz, consonant, vowel, and sentence recognition improved as a function of the number of channels and reached plateau at 4 to 6 channels. Subjective judgments of sound quality continued to improve as the number of channels increased to 12, the highest number tested. Tone recognition, i.e., recognition of the four Mandarin tone patterns, depended on both the number of channels and the low-pass cutoff frequency. The trade-off between the temporal and spectral cues for tone recognition indicates that temporal cues can compensate for diminished spectral cues for tone recognition and vice versa. An additional tone recognition experiment using syllables of equal duration showed a marked decrease in performance, indicating that duration cues contribute to tone recognition. A third experiment showed that recognition of processed FM patterns that mimic Mandarin tone patterns was poor when temporal envelope and duration cues were removed.

I. INTRODUCTION

More than a quarter of the world's population speak one of the tone languages, which include Mandarin Chinese, Cantonese, Vietnamese, Thai, etc. In fact, Mandarin Chinese is spoken by more people than any other language. In Mandarin Chinese, there are four tone patterns as defined by the fundamental frequency of voiced speech. Other features that vary with the tone patterns include amplitude pattern and syllable duration. In isolated syllables, the patterns of the fundamental frequency for tone patterns, referred to as tone 1, 2, 3, and 4 in the following, are (1) flat and high; (2) rising; (3) falling and then rising; and (4) falling [Fig. 1(A)]. The distinguishing characteristic of the tone language is that the variations of the tone patterns of each syllable convey lexical meaning. For example, tones 1 through 4 associated with the Chinese syllable /xu/ (pronounced as “shoo”) can mean (1) “void;” (2) “slowly;” (3) “permit;” and (4) “sequence;” respectively.

Correspondence to: Li Xu.

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

The contemporary cochlear implant has provided benefit to thousands of patients with profound hearing loss throughout the world. However, for socioeconomic reasons, only a relatively small number of deaf people in China and Southeast Asia, where tone languages are most prevalent, have received the commercially available cochlear implants. Sporadic reports with a limited number of native Chinese speaking patients have shown that those patients, using the multi-channel cochlear implants, have achieved tone recognition at various levels. Tone recognition varies from around chance performance to nearly 100% correct levels (Xu *et al.*, 1987; Zeng, 1995; Huang *et al.*, 1996; Sun *et al.*, 1998; Wei *et al.*, 2000).

What features of the tone languages are important for tonal-speech recognition? The answer to this question is crucial for understanding the mechanism by which the implant patients perceive the tone patterns of tone languages and for designing future cochlear implants that are effective for patients who speak tone languages. In the current speech processing strategies of cochlear implants, the speech signal is typically represented in electrical pulses that are amplitude modulated by the envelope of the signal. In the CIS (continuous interleaved sampling) strategy (Wilson *et al.*, 1991), the speech signal is divided into a number of bands and the envelope of each band is extracted and used to modulate the pulse trains. In the SPEAK (spectral peak) or ACE (advanced combined encoding) strategies (Skinner *et al.*, 1994; Seligman and McDermott, 1995; Cochlear Corporation, 1999), the speech signal is sent to a bank of filters (20 for SPEAK and 22 for ACE) and the envelopes of five to ten of the largest outputs of the 20 or 22 filters are extracted. In a recent study, Fu and colleagues (1998b) simulated one-through four-channel CIS-type cochlear prosthesis and tested speech recognition in normal-hearing native Chinese speakers. Their results showed that the recognition scores of Chinese vowels, consonants, and sentences bore a similar pattern to that seen in English (Shannon *et al.*, 1995). They also found that tone recognition scores were high and largely independent of the number of filter bands. In contrast, Wei and colleagues (2001) showed in a preliminary report that the tone-recognition performance in a group of four cochlear implant users depended on the number of channels and reached a plateau at about ten channels. These apparently conflicting results motivated the following questions: (1) what would the tonal-speech recognition performance be with more than four channels; (2) how would the subjects judge the sound quality of the acoustic simulations of cochlear prosthesis; and (3) what features of the speech signals provide the robust cues for tone recognition under cochlear prosthesis-type stimulation?

The present study was designed to address those questions. Using the methods of acoustic simulations of cochlear prosthesis as developed and used by Shannon and colleagues (1995, 1998) and other researchers (Dorman *et al.*, 1997a, 1997b), we examined the features of stimulation that affect tonal-speech recognition and subjective sound quality judgments. The effect of spectral resolution was studied by varying the number of simulated channels from 1 to 12. The temporal features of the speech were studied by varying the low-pass cutoff frequencies used in extracting the amplitude envelopes. According to Rosen (1992), the temporal features of speech signals can be partitioned into three categories, i.e., envelope (2-50 Hz), periodicity (50-500 Hz), and fine structure (500-10 000 Hz). In the present study we modified the envelope and periodicity cues as the temporal features for speech recognition of Mandarin Chinese. We varied the low-pass cutoff frequencies systematically from 1 to 512 Hz in 1-octave steps, to examine the details of the temporal features of stimulation that might contribute to the tone recognition. The effects on tone recognition of the interaction between the temporal and spectral features of stimulation were then studied. In order to gain control of the syllable duration and overall amplitude (i.e., suprasegmental features), which might convey information for tone recognition, new sets of speech signals that were carefully matched in duration were recorded. Tone recognition on those equal-duration syllables was tested at randomly varying sound levels. In addition, frequency-modulated (FM) sweeps that mimicked the Mandarin Chinese tone patterns were generated [Fig. 1(B)] so as to gain parametric control

of the temporal and spectral features of the stimuli. Recognition of the FM sweeps processed through the cochlear prosthesis simulation models was then tested.

II. METHOD

A. Subjects

Ten normal-hearing native Mandarin Chinese speakers were recruited from the student and staff population at the University of Michigan. All subjects had pure-tone air-conduction thresholds ≤ 20 dB HL at octave frequencies between 250 and 8000 Hz. Subjects ranged from 29 to 43 years of age (35.6 ± 64.4 , mean and s.d.). There were eight females and two males. All subjects had at least college-level education in China. The use of human subjects in this study was reviewed and approved by the University of Michigan Medical School Institutional Review Board.

B. Speech test materials and frequency-modulation sweeps

The Mandarin Chinese speech test materials consisted of consonant, vowel, tone, and sentence tests in both male and female voices. All of the combinations of consonant-vowel-tone syllables used in this study were real words in Chinese. The tone, consonant, and vowel tests were in a four-choice format in which the subjects were required to choose one of the selections from a list of four words. The consonant and vowel lists of the male and female voices were compiled from the Speech Audiometry Tables developed by the Department of Otolaryngology at the PLA General Hospital, Beijing (Gu, 1988). There were 40 lists of four Chinese monosyllable words for consonant and vowel tests. In each list for the consonant test, the four syllables had the same vowel and tone, e.g., du(2), mu(2), ru(2), and shu(2). In each list for the vowel test, the four syllables had the same consonant and tone, e.g., da(4), dai(4), di(4), and du(4). The tone and sentence tests of the male voice were initially developed by Beijing Union Hospital for audiometric test of patients (Zhang *et al.*, 1988) and were further compiled by Fu *et al.* (1998b). We recorded the female voice counterpart of the tone and sentence test materials. There were ten lists of four syllables for the tone tests and each syllable in a given list had the same consonant and vowel, e.g., ma(1), ma(2), ma(3), and ma(4). For each tone test, the ten lists were administered four times for a total of 40 lists. In these 40 lists, the male voice appeared in half of the cases and the female voice appeared in the other half, in random order. The sentence test materials consisted of ten lists of ten short sentences, each of which had four to seven key words. Each sentence test used one of the sentence lists. Additionally, we recorded 50 different short sentences used for practice purposes only. All speech test materials were digitized at a sampling rate of 22 050 Hz and stored in a 16-bit format.

The duration of naturally spoken monosyllable words varied systematically with the tone patterns, with tone 3 being the longest and tone 4 the shortest (Whalen and Xu, 1992; also see Fig. 8). This duration cue might contribute to the recognition of the tones. To eliminate this potential cue, we conducted an additional experiment in which we re-recorded the ten lists of the tone test materials multiple times using both male and female voices. The waveforms of the multiple recordings were then visualized with sound-analysis software [COOLEDIT 2000 (Syntrillium Software, Scottsdale, AZ)]. Tokens in which the durations of all four tones of the same syllable were within 1-ms precision were selected. These equal-duration speech materials were then used for additional tone recognition tests as described below.

To gain better parametric control of the stimuli, two sets of four frequency-modulation (FM) sweeps that mimicked the four Mandarin Chinese tone patterns were generated for recognition tests as described below. All FM sweeps were pulse trains of 500 ms in duration. Table I lists the initial, middle, and final fundamental frequencies (F_0) of the two sets of four patterns of FM sweeps. The sampling rate was of 30 kHz. The F_0 of the higher-pitched FM sweeps were

1 octave above the lower-pitched FM sweeps. The F_0 of the two sets of FM sweeps were chosen to represent the upper range of the F_0 of a male and a female voice (Fletcher, 1953; Lehiste, 1970). Different patterns and sets of FM sweeps were achieved by varying the interpulse intervals of the pulse train. Frequency analysis of the FM sweeps confirmed that they contained the F_0 and all the high harmonics of equal energy up to the Nyquist frequency (15 kHz). Figure 1(B) shows the spectrograms of the higher-pitched FM patterns.

C. Signal processing

Signal processing for acoustic simulations of cochlear prosthesis was performed in MATLAB (MathWorks, Natick, MA). Mandarin Chinese speech signals were first passed through a bank of analysis filters (sixth-order elliptic bandpass filters). To study the effects of number of channels on recognition of tones, consonants, vowels, and sentences, the number of analysis filters was varied between 1 and 12 (1, 2, 3, 4, 6, 8, 10, and 12). The overall bandwidth was from 150 to 5500 Hz. The bandwidth of each analysis filter was based on estimated equal distance along the basilar membrane of the cochlea. The corner frequencies (F) were calculated with Greenwood's (1990) formula, $F = 165.4(10^{0.06x} - 1)$, where x is the distance in mm from the apex. For instance, the corner frequencies for the 6-channel analysis filters were 150, 345, 660, 1171, 1997, 3335, and 5500 Hz. Next, the temporal envelope of each analysis band was extracted by half-wave rectification and low-pass filtering (second-order Butterworth filters). The low-pass cutoff frequencies (LPFs) of the low-pass filters were fixed at 512 Hz for speech materials of consonants, vowels, and sentences, but were varied between 1 and 512 Hz in 1-octave steps for the tones in order to study the effects of the temporal features on the tone recognition. The temporal envelope of each band was then used to modulate a white noise that was bandpassed through the same analysis filter. Finally, the modulated noise bands were summed and the resultant signals were stored on the computer hard disk for presentation.

For the FM sweep recognition, FM signals were processed through the same cochlear prosthesis simulation models with number of channels varying from 4 to 40 (4, 6, 8, 10, 12, 16, 20, 30, and 40), and the LPFs varying in 2-octave steps from 4 to 1024 Hz (4, 16, 64, 256, and 1024 Hz). In pilot experiments, we found that the higher numbers of channels and higher LPFs were required to achieve recognition more than twice the chance performance with the FM sweeps.

D. Procedures

The speech signals were presented through a loudspeaker positioned 1 m away from the subject at 0° azimuth inside an Acoustic Systems (model RE2 242S) double-walled sound-attenuating booth. The subjects were free to adjust the sound volume to the most comfortable levels; levels typically measured ~ 60 dB (A) with a sound level meter (Brüel and Kjær type 2231). A graphical user interface (GUI) was built to present the tone, consonant, and vowel tests in the four-choice format. In the GUI, four buttons labeled with pinyin (i.e., phonemic spellings) and Chinese characters were shown on a computer screen. After listening to the acoustic presentation of each speech signal, the subjects were required to indicate what they had heard by pointing and clicking on one of the four buttons using a computer mouse. After each response, the GUI would refresh the screen, display a new list of four choices, and present the next stimulus. A tone, consonant, or vowel test consisted of 40 four-choice trials. The sentence test consisted of ten sentences. The subjects were required to write down each sentence that they had heard. A different GUI was built to record the subjective judgments of the sound quality. In this GUI, a sliding bar was plotted from the left side of the screen to the right side of the screen, representing poor to excellent sound quality. The subjects were instructed to slide a marker, using a computer mouse, to a location that represented their judgment of the sound quality of the speech signals ("very poor" to "excellent"). For the FM pattern recognition test, the GUI showed four buttons labeled with the four respective patterns of the FM [(1) -, (2) /, (3) V, and (4) \].

All subjects received 4 to 5 h of training in listening to the processed speech materials. The speech materials for training consisted of 20 sentences randomly selected from the 50 practice sentences for all the testing conditions (i.e., number of channels) starting with 12 channels and ending with 1 channel. Subjects reported verbally what they heard. Feedback was provided during practice. Additional brief training was provided for consonant, vowel, and tone recognition tests to familiarize the subjects with the task and the use of a computer mouse.

The experiments were divided into four sessions. In the first session, subjects were tested on consonant, vowel, tone, and sentence recognition of speech materials that were processed through 1 to 12 channels with the LPF fixed at 512 Hz. The testing order for the number of channels was randomized. Each subject performed two speech tests that used both male and female voices that were presented separately. Thus, a total of 160 items was tested for consonant, vowel, and tone recognition and 40 sentences (about 200 words) were tested for sentence recognition for each number of channels. After each set of consonant, vowel, tone, and sentence tests, a subjective judgment of the sound quality was obtained. One of the subjects (Subject 1) was the female speaker of the sentence materials. Therefore, she was not tested on sentence recognition and no subjective judgments of the sound quality were obtained from her.

In the second session, subjects were tested for tone recognition only. The signals were processed through 1 to 12 analysis channels with the LPFs varied from 1 to 512 Hz in 1-octave steps. Subjects were tested twice on male voice and twice on female voice (a total of 160 items) for each combination of number of channels and LPFs. Subject 6 could not participate in this experiment session.

The third session was similar to the second session except that the speech signals were of equal durations for the four tone patterns. Also, the level of each presentation was randomized among 55, 60, 65, and 70 dB (A) to confound possible cues derived from the overall amplitude. Four of the ten subjects (Subject 2, 3, 7, and 8) participated in this session of the experiment.

The fourth session tested the recognition of the four FM sweeps that were processed through the cochlear prosthesis model. The subjects were trained for about half an hour on the unprocessed and processed FM sweeps. A total of 7200 (i.e., $80 \times 9 \times 5 \times 2$) responses was collected from each subject to 80 presentations of each combination of number of channels (9) and LPFs (5) and pitch (2). The order of stimuli was randomized. The level of each presentation was also randomized as in the third session. The same four of the ten subjects (Subject 2, 3, 7, and 8) who participated in the third session participated in this session of the experiment.

E. Data analysis

Arcsine-transformed percent-correct data from the speech recognition tests were used to compare the speech recognition performance under various test conditions. The arcsine transformation takes the form of $y = 2 \times \sin^{-1} \sqrt{p}$, where p is a percent-correct score. The rationale for arcsine transformation is that the data of percent-correct have nonuniform variance, whereas the transformed data have the property of stabilized variance of binomial data and thus are more suitable for analysis of variance (ANOVA) and other statistical analysis (Thornton and Raffin, 1978; Studebaker, 1985). A one-way ANOVA was employed to determine whether the effects on speech recognition of the independent variables (number of channels or LPFs) were statistically significant. A two-way ANOVA was used to determine whether tone recognition scores were affected by the number of channels or LPFs or both. When the ANOVA revealed a significant difference, the Tukey test was used for post hoc pairwise comparisons of the speech-recognition scores among the dependent variables such as number of channels or LPFs. The Tukey test was chosen because it provides a good control of the increased cumulative type-I errors inherent in repeated pairwise comparison (Keppel, 1991).

To study the tone recognition based on the syllable duration alone, a maximum-likelihood estimator was used as an ideal observer (Green and Swets, 1966). In that procedure, four histograms were generated for the durations of four tone patterns using a time bin of 60 ms. The number of syllables that fell into each time bin was then compared across the four histograms. The maximum numbers of all of the time bins were then summed and the sum was divided by the total number of syllables to derive a percent-correct value for the ideal observer.

III. RESULTS

We found no significant differences between speech recognition scores for male versus female voices (t test, $p > 0.05$). The frequency analysis of the male and female voices showed that the fundamental frequencies for the flat tone pattern (tone 1) were approximately 150 and 300 Hz, respectively. It seemed that the periodicity features did not contribute significantly to Mandarin Chinese speech recognition using the cochlear prosthesis acoustic simulations or that the effects of the periodicity features were confounded by other features in the speech signal. In the following report, the speech recognition scores of both male and female voices were combined.

A. Dependence on number of channels

Figure 2 (upper panels) summarizes performance in tone, consonant, vowel, and sentence recognition as a function of number of channels. The speech recognition scores (upper panels) increased as the number of channels increased from one to six and plateaued at the number of channels greater than or equal to six. For number of channels below six, the strength of dependence on number of channels was larger for vowel recognition than for consonant recognition. The strength of dependence on number of channels was the weakest for tone recognition. The dependence of sentence recognition on the number of channels between one and six was fairly strong, although this dependence was not readily comparable to the dependence on number of channels for tone, consonant, and vowel recognition because the tests had different chance levels to start with. It is noteworthy that the mean performance of sentence recognition with only one channel was a little above 10% correct, consistent with the results of the Fu *et al.* (1998b) study. One of our subjects reached a surprisingly high score of 35% correct. For native English speakers, the one-channel acoustic simulations resulted in an average of only 3% correct for English sentence recognition (Shannon *et al.*, 1995). The apparent differences between Mandarin Chinese and English suggest that the high-level tone recognition of Mandarin Chinese might contribute to the relatively higher sentence-recognition scores under the situations of one-channel stimulation.

The ANOVA revealed that the mean speech recognition scores were significantly different across numbers of channels ($p < 0.01$). Pairwise comparisons between the various numbers of channels were then carried out for speech recognition scores. The results are shown in the lower panels of Fig. 2, in which each small square represents one comparison, and light- and dark-gray-filled squares represent significance levels of $p < 0.05$ and $p < 0.01$ (the Tukey test), respectively. The tone-recognition scores with the number of channels greater than or equal to three were significantly higher than those with only one channel. Also, the tone-recognition scores with 8 and 12 channels were significantly higher than those with two channels. Beside those differences, a larger number of channels in general did not result in significantly higher scores for tone recognition. The consonant-, vowel-, and sentence-recognition scores increased as a function of number of channels from one through six. While no further improvement was evident for consonant and sentence recognition with number of channels beyond six, vowel recognition did show small but significant improvement with a larger number of channels.

B. Subjective judgments

Despite the fact that the speech-recognition scores generally reached a plateau at six channels, the subjective judgments of sound quality continued to favor a larger number of channels. Figure 3 plots the distribution of mean quality judgments as a function of number of channels obtained from nine of the ten subjects. The range of quality judgments was highly variable from subject to subject. The quality judgments across all speech tests ranged from 0.10 to 0.99 for one subject to 0.02 to 0.64 for another subject. To facilitate comparison across subjects, all quality judgments from each subject were normalized to his/her maximum quality judgment of all tests. The normalized quality judgments of each condition (i.e., number of channels) were then averaged. Among subjects, the normalized quality judgments showed large variability. A side effect of this normalization procedure was that it minimized the differences in absolute levels of quality judgments among subjects. Nevertheless, there was a general trend that subjective sound quality improved as the number of channels increased up to 12.

Figure 4 plots the relationships between the speech recognition scores for tones, consonants, vowels, and sentences and the normalized quality judgments. Each dot represents results from one administration of the speech tests for one subject. The correlation coefficients (r) were fairly high for the vowel and sentence perception, but low or moderate for the tone and consonant perception. It was not surprising to see that the tone recognition scores and the quality judgments showed weak correlation because the tone recognition exhibited little dependence on number of channels (Fig. 2), whereas the subjective judgments of sound quality generally depended on the number of channels (Fig. 3). This result was also in agreement with Lehiste's (1970) observations in other tone languages that subjective judgments of phonetic quality did not seem to depend on tone recognition. It was interesting that the vowel and sentence recognition correlated more strongly with the subjective judgment of sound quality than did the consonant recognition. Perhaps the voicing of the speech contributed more to the subjective judgments than did the unvoiced part of the speech.

C. Dependence of tone recognition on the temporal features of stimulation

The low-pass cutoff frequencies (LPFs) of the low-pass filters used in extracting the amplitude envelopes determine the amount of temporal detail in the envelope of the signal in each channel. Fu and colleagues (1998b) tested the tone recognition at LPFs of 50 and 500 Hz for one- through four-channel processors and found that the tone recognition scores were around 65% and 80%, respectively. In the present study, we examined the dependence of tone recognition on the LPFs that ranged from 1 to 512 Hz in 1-octave steps. Figure 5 shows the distribution of the pooled mean percent correct for tone recognition using 1 or 12 channels. Using 12 channels (Fig. 5, upper trace) typically resulted in better performance of tone recognition than using only 1 channel (Fig. 5, lower trace), as shown also in Fig. 2. The increase in the tone recognition scores as a function of LPFs was consistent. The ANOVA showed highly significant differences in the tone recognition scores among LPFs for all the channel numbers ($p < 0.01$).

D. Interaction between spectral and temporal cues for tone recognition

Tone recognition was tested in nine subjects on all combinations of eight different numbers of channels and ten low-pass cutoff frequencies (LPFs) of the filter used for the envelope extraction. Figure 6 shows the percent-correct results for tone recognition from three individual subjects. Subject 3 [Fig. 6(A)] had relatively low tone recognition scores compared to other subjects. In contrast, Subject 5 [Fig. 6(C)] had relatively high scores. Subject 4 [Fig. 6(B)], however, was representative of most subjects. Most subjects showed high tone recognition scores with larger number of channels and/or with higher LPFs. There was a gradient of performance along the main diagonal line in plots like Fig. 6. This illustrated a trade-off between the number of channels and LPFs for the tone recognition. Using a small number of

channels combined with a high LPF, subjects could achieve a tone recognition performance similar to that achieved using a large number of channels but a low LPF.

The pooled results from the nine subjects who participated in this session (session two) of the experiments are shown in Fig. 7(A). Note the trade-off between the number of channels and the LPFs on the tone recognition performance, similar to the patterns seen for individual subjects in Fig. 6. Two-way ANOVA revealed that both number of channels and LPF affected the tone recognition scores ($p < 0.01$). The pooled data also indicated that the tone recognition performance varied little with number of channels in most cases when the number of channels was greater than six. Similarly, the tone recognition performance was not affected by LPFs when the LPFs were ≥ 128 Hz. High levels of performance of tone recognition could only be achieved with either high LPFs or a large number of channels or both. The low levels of performance with LPFs equal to 1 or 2 Hz were around 40 to 50 percent correct, which was still higher than the chance performance of 25%. This result suggested that the tone information was very robust in the speech materials or that subjects were using some additional cue to help categorize the tokens.

E. Tone recognition without syllable duration as a potential cue

The durations of Mandarin Chinese syllables vary systematically with the tone patterns, and this could contribute to categorization of the tokens. Figure 8 plots the syllable durations of tone patterns 1 through 4 for speech test materials used in session 2, reported above. There were ten syllables for male voice and ten syllables for female voice for each tone pattern. In general, tone 3 was the longest and tone 4 was the shortest in duration; tone 2 tended to be slightly longer than tone 1. The maximum-likelihood estimator for tone recognition based on the duration alone yielded percent correct of 52.5% for male voice, 62.5% for female voice, and 56.5% when both male and female voices were combined. Note that these levels of performance were only slightly higher than those observed with one-channel processors where the subjects were forced to use only duration and amplitude cues.

In session 3, using syllables of equal duration for the four tones in the tone recognition test, we found a large decrease in the tone recognition performance for all of the combinations of number of channels and LPFs. Figure 7(B) shows the pooled results from the four subjects who participated in this session of the experiments. Compared to Fig. 7(A), the decreases in performance ranged between 6.0 to 24.9 percentage points with a mean of 15.1 percentage points. Across all tone recognition tests, the decreases in performance were 12.5, 8.5, 19.5, and 19.7 percentage points for tone 1 through 4, respectively. Tone 3 and tone 4 were the longest and shortest in duration (Fig. 8) and removal of the duration cue affected the performance for these two tones the most. These results indicate that the suprasegmental features such as duration of the syllables contributed substantially to the high-level performance of the tone recognition in session 2. Despite the decrease in performance in general when the duration cues were removed, the trade-off between the number of channels and the LPFs was still present (two-way ANOVA, $p < 0.01$). For example, to achieve 56.25% correct of tone recognition with one channel, one would need to have $\text{LPF} > 256$ Hz; however, to achieve the same level of performance with 12 channels, one could use LPF as low as 4 Hz. The other difference between Figs. 7(A) and 7(B) was the ranges of number of channels and LPF on which tone recognition depended. In Fig. 7(A), tone recognition performance increased as number of channels increased from one to six and as LPFs increased from 1 to 128 Hz. In Fig. 7(B), the performance increased as number of channels increased from one to 12 and as LPFs increased from 1 to 256 Hz.

F. Recognition of FM sweeps processed through the cochlear prosthesis simulations

FM sweeps provided an ideal control for duration and temporal envelope, when testing recognition of tone patterns (Fig. 1). It was not surprising that the recognition of the processed FM sweeps was more difficult for the subjects than was recognition of the speech material, because no temporal envelope cues or suprasegmental cues were available to the listeners. Figure 9 shows the mean performance of the four subjects who participated in this session of the experiment. Panels (A) and (B) represent recognition scores for the lower- and higher-pitched FM sweeps processed through the cochlear prosthesis simulation models. Panel (C) represents the pooled data for both lower- and higher-pitched FM sweeps. For the lower-pitched FM sweeps, relatively high performance was achieved at the low-pass cutoff frequency (LPF) of 64 Hz and above [Fig. 9(A)] and relatively little effect of the number of channels were seen. Relatively high performance for the higher-pitched FM sweeps required the LPF of 256 Hz and above [Fig. 9(B)]. In general, little dependency of the performance was found on the number of channels of up to 20. When the number of channels were as high as 30 and 40, however, a remarkable improvement of the recognition performance was noted for the higher-pitched FM sweeps at almost all LPFs compared to the performances for the lower-pitched FM. The trade-off of the LPF and the number of channels for the recognition performance was weak at best for the range of LPFs and number of channels tested.

IV. DISCUSSION

In normal-hearing subjects listening to acoustic simulations of cochlear prosthesis, four to six independent channels of information are necessary and sufficient to achieve high-level English speech recognition under ideal situations (Shannon *et al.*, 1995; Dorman *et al.*, 1997a; Loizou *et al.*, 1999). Fu and colleagues (1998b) tested speech recognition for Mandarin Chinese in a group of native Mandarin Chinese speakers presented with one- through four-channel acoustic simulations of cochlear prosthesis. The subjects achieved similar consonant, vowel, and sentence recognition to the English counterparts. The tone recognition, however, was found to be independent of the number of channels. The present study extended those observations to a much greater range of channels (i.e., from 1 to 12). Our results were, in general, consistent with previous observations that the Chinese consonant, vowel, and sentence recognition improved as a function of the number of channels. We also demonstrated the dependence of tone recognition on number of channels, which is discussed below. We found that six channels were necessary for the consonant, vowel, and sentence recognition in quiet to reach optimum performance. Beyond six channels, the vowel recognition improved slightly, but by and large, the consonant-, vowel-, and sentence-recognition performance reached a plateau (Fig. 2).

It should be noted that under more difficult listening conditions, such as in the presence of background noise, the number of channels needed for high-level English speech recognition was found to be much larger, depending on the signal/noise ratio (Dorman *et al.*, 1998; Fu *et al.*, 1998a; Friesen *et al.*, 2001).

There is evidence that cochlear prosthesis can utilize only four to six effective channels (Fishman *et al.*, 1997; Dorman and Loizou, 1998; Fu *et al.*, 1998a; Friesen *et al.*, 2001). In addition, findings from studies using acoustic simulations of cochlear prosthesis in normal-hearing subjects and those from studies in cochlear implant patients are in agreement in that optimum speech recognition in quiet can be achieved with a few channels. This clearly has practical importance when we design an effective, low-cost implant for the developing countries including China (Zeng, 1995; Wilson *et al.*, 1998). We found that high scores of tone recognition with a few channels were achieved using the speech materials in which the syllable duration was not equalized for the four tone patterns. However, syllable duration clearly contributed to the relatively high performance in tone recognition (Fig. 8). Studies using information transmission analysis showed that implant patients were able to utilize duration

cues in consonant (van Wieringen and Wouters, 1999) and vowel (Tyler *et al.*, 1992; van Wieringen and Wouters, 1999) recognition when consonants and vowels were presented in isolation. In everyday speech, syllable duration is not a reliable cue for speech recognition. Different speakers might talk at different speeds. The same speaker might change speed depending on many factors, such as the emotional state of the speaker or the context of the sentences. While listeners might adapt to these changes to some extent, variability in talker speed is likely to weaken the duration cues. When we tested the tone recognition using syllables of equal duration for the four tone patterns to eliminate duration as a cue, we found the tone recognition scores dropped an average of about 15 percentage points. Such a decrease in performance was observed within the range of numbers of channels (i.e., 1 to 12) that we tested (Fig. 7).

The best tone recognition scores with 12 channels and a 512-Hz LPF when duration cues were removed were only about 75% correct. In a preliminary report, Wei and colleagues (2001) found that tone recognition of the best performer of a group of four native Chinese speaking patients with cochlear implants was about 80% correct with the number of channels being 7 or 10. Thus, there is room for improvement in tone perception by the cochlear implant users. Current speech coding strategies are not designed to optimize the tone perception. Improvement of pitch perception for implant users is currently an active research goal (e.g., Faulkner *et al.*, 2000; Geurts and Wouters, 2001). That research is applicable to tonal speech and should profit from research done in the tone language speakers.

Although subjective sound quality is important to implant patients, few studies in the field of cochlear implant research have systematically examined the subjective judgments of the sound quality of electric hearing in patients or acoustic hearing of simulations of cochlear implants in normal hearing subjects. Pfingst and colleagues (2001) indicated that the subjective judgments of the quality were only moderately correlated with the subject's speech recognition ability, which suggested that the mechanisms contributing to sound quality and speech recognition ability do not completely overlap. Fishman and colleagues (1997) evaluated the subjective benefit in implant patients of experimental processors that varied in the number of channels. They found that the subjective judgments of the benefit with number of channels up to 10 exceeded that with fewer channels. Little further benefit was seen with 20 channels compared to 10 channels. In the present study we demonstrated that the normal hearing subjects' preferences of the sound quality increased as a function of the simulated number of channels up to 12, the highest number tested (Fig. 3). Caution should be exerted, however, when extrapolating the results from normal-hearing subjects to implant patients. One of the reasons is that we do not know if it is possible to activate a large number of *independent* channels in current cochlear prostheses. Functional channel overlap might negate the potential benefit of a large number of channels. Nevertheless, the present study indicates that even though the speech recognition scores in quiet have reached a plateau at four to six channels, more independent channels can still bring about improvement of the sound quality.

Tone patterns of tone languages are defined by patterns of the fundamental frequency (F_0) for the voiced part of the word (see Fig. 1). Yet, F_0 itself does not seem to be essential for tone pattern recognition. Liang (1963) tested the tone recognition of high-passed Mandarin Chinese speech signals. He found that high-level tone recognition was preserved for signals that were high-passed above 2.4 kHz. One might attribute the tone recognition to the residue pitch (Schouten *et al.*, 1962). In another experiment, however, Liang (1963) found that the tone recognition maintained at 60- to 70-percent correct for whispered speech in which neither F_0 nor its harmonics existed. Whalen and Xu (1992) pointed out that the Mandarin tone information might be carried by the amplitude contour (temporal envelope) of the speech signals. Their subjects maintained high-level recognition of tones when the natural speech was transformed to signal-correlated noise (Schroeder, 1968) in which no F_0 or formant structure

existed but the original amplitude contour was retained. Fu and colleagues (1998b) found that tones were consistently recognized at about the 80%-correct level using acoustic simulations of one- to four-channel CIS-type processing strategies. These results supported the hypothesis that the temporal envelope carries information about the tone patterns. Therefore, as discussed below, under situations of restricted spectral information as in cochlear prostheses, faithful transmission of temporal envelope information would be important for tone recognition.

The present study confirmed the results of Fu *et al.* (1998b) in that high-level recognition of tones could be achieved with a small number of channels and that the more robust temporal information in the envelope (i.e., LPF of 500 Hz versus 50 Hz) was beneficial for tone recognition. Fu and colleagues (1998b) indicated that the tone recognition was not dependent on the number of channels (i.e., one through four). However, our results demonstrated that the tone recognition depended on the number of channels, although the dependence was not as strong as that for vowel or consonant recognition (Fig. 2). In the Fu *et al.* (1998b) study, only a male voice was used for the tone recognition test. In the present study, both male and female voices were used and thus the test became more difficult for the subjects, so that their performance with one or two channels did not reach as high a level as with more channels. When the syllable duration was equalized, the tone recognition became even more challenging and the dependence on the number of channels was more apparent (Fig. 7). Results from cochlear implant users support the conclusion that tone recognition is affected by the number of channels (Wei *et al.*, 2001).

When the spectral information is minimized, the temporal information of the speech signal has been found to contribute to consonant recognition, especially for voicing and manner (Van Tasell *et al.*, 1987; ter Keurs *et al.*, 1992, 1993; Drullman *et al.*, 1994) and for certain aspects of vowel recognition, e.g., short versus long vowels (Rosen, 1992; Drullman *et al.*, 1994). However, little is known about the effects of interaction between spectral cues and temporal cues on speech recognition. For tone recognition, we found that there was a trade-off between temporal and spectral information. The trade-off existed between the number of channels, ranging from one to six or eight, and the LPFs, ranging from 1 to 128 Hz [Fig. 7(A)]. Beyond these, the tone recognition became insensitive to either an increase in number of channels or an increase in the LPFs. The higher LPFs (i.e., 256 and 512 Hz) should permit some or all periodicity information to pass through the acoustic simulations, yet the tone recognition showed no more improvement. This might be due to the already high-level recognition at LPF of 128 Hz leaving little room for improvement. Another interesting observation was that the tone recognition with LPFs of 1 or 2 Hz was 40- to 50 percent correct [Fig. 7(A)], nearly twice as high as the chance performance of 25 percent correct. This suggests that the suprasegmental features of the speech signal (Lehiste, 1970), such as duration of the syllables, probably contribute to the recognition among tone patterns. We specifically addressed this issue by using equal-duration syllables for the tone recognition tests. Under these conditions, we found that the trade-off between the temporal and spectral information for tone recognition extended up to 12 channels and a LPF of 256 Hz [Fig. 7(B)]. Such a trade-off between the number of channels and the LPFs essentially disappeared when the subjects were tested on the recognition of the processed FM sweep patterns (Fig. 9). This is not surprising because there is no temporal envelope or amplitude contour information in those signals. The high LPFs (64 Hz for the lower-pitched FM sweeps and 256 Hz for the higher-pitched FM sweeps) permit the periodicity information to pass through the simulation. In fact, the subjects reported that they could hear the buzz of pulse trains when the LPFs were above the F_0 of the FM sweeps. When the LPFs were near the F_0 of FM patterns, the subjects could perform the task by listening to a transition of a buzz sound to a hiss sound (pattern 2) or a transition of a hiss sound to buzz sound (pattern 4). The improved recognition at 30 and 40 channels was probably due to spectral information that was made available to the listeners through the simulations. Since the frequency spacing of the harmonics of the higher-pitched FM sweeps are twice as large as the spacing of the

harmonics of the lower-pitched FM sweeps [Fig. 1(B)], simulations using 30 to 40 channels resolve some of the spectral cues for the higher-pitched FM sweeps but much less for the lower-pitched FM sweeps.

V. CONCLUSIONS

With acoustic simulation of a CIS processor, high levels of speech recognition of Mandarin Chinese in quiet could be achieved with four to six channels, although the sound quality improved further with more channels. Good tone recognition was also achieved with a few channels given robust temporal information (e.g., LPFs > 100 Hz). However, the high-level tone recognition might have been due in part to suprasegmental cues (such as syllable duration) in the speech materials. When we controlled for those suprasegmental cues, the tone-recognition performance dropped markedly. Tone recognition depended on both the number of channels and the LPFs, and such dependency was reflected in a tradeoff between the spectral and temporal cues for the tone recognition. When the temporal envelope cue was removed, as in the case of recognition of processed FM patterns, performance drastically deteriorated and in such a situation, spectral information presented in large numbers of channels (≥ 30) and temporal information presented in high LPF (>64 Hz) could be useful for tone recognition.

ACKNOWLEDGMENTS

We are grateful to Qianjie Fu and Xiaosong Wang for providing some of the speech testing materials used in this study. John Middlebrooks, San Duanmu, and Kevin Franck provided valuable comments on an earlier version of the manuscript. We also thank Bob Shannon and an anonymous reviewer for their constructive comments. We acknowledge the technical assistance from Jianzhong Lu and Julie Oldfield. The work was supported by NIH/NIDCD Grant Nos. F32 DC00470, R01 DC03808, and T32 DC00011.

References

- Cochlear Corporation. Lane Cove. Cochlear Limited; Australia: 1999. Nucleus Technical Reference Manual Z43470 Issue 1.
- Dorman MF, Loizou PC. "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels,". *Ear Hear* 1998;19:162–166. [PubMed: 9562538]
- Dorman MF, Loizou PC, Fitzke J, Tu Z. "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels,". *J. Acoust. Soc. Am* 1998;104:3583–3596. [PubMed: 9857516]
- Dorman MF, Loizou PC, Rainey D. "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise outputs,". *J. Acoust. Soc. Am* 1997a;102:2403–2411. [PubMed: 9348698]
- Dorman MF, Loizou PC, Rainey D. "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,". *J. Acoust. Soc. Am* 1997b;102:2993–2996. [PubMed: 9373986]
- Drullman R, Festen JM, Plomp R. "Effect of temporal envelope smearing on speech reception,". *J. Acoust. Soc. Am* 1994;95:1053–1064. [PubMed: 8132899]
- Faulkner A, Rosen S, Smith C. "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vo-coded speech: Implications for cochlear implants,". *J. Acoust. Soc. Am* 2000;108:1877–1887. [PubMed: 11051514]
- Fishman KE, Shannon RV, Slattery WH. "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,". *J. Speech Lang. Hear. Res* 1997;40:1201–1215. [PubMed: 9328890]
- Fletcher, H. *Speech and Hearing in Communication*. Van Nostrand; Princeton, NJ: 1953.
- Friesen L, Shannon RV, Baskent D, Wang X. "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,". *J. Acoust. Soc. Am* 2001;110:1150–1163. [PubMed: 11519582]

- Fu QJ, Shannon RV, Wang X. "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,". *J. Acoust. Soc. Am* 1998a;104:3586–3596. [PubMed: 9857517]
- Fu QJ, Zeng FG, Shannon RV, Soli SD. "Importance of tonal envelope cues in Chinese speech recognition,". *J. Acoust. Soc. Am* 1998b;104:505–510. [PubMed: 9670541]
- Geurts L, Wouters J. "Coding of the fundamental frequency in continuous interleaved sampling for cochlear implants,". *J. Acoust. Soc. Am* 2001;109:713–726. [PubMed: 11248975]
- Green, DM.; Swets, JA. *Signal Detection Theory and Psychophysics*. Wiley; New York: 1966.
- Greenwood DD. "A cochlear frequency-position function for several species' 29 years later,". *J. Acoust. Soc. Am* 1990;87:2592–2605. [PubMed: 2373794]
- Gu R. "Staggered spondaic word test and the competing sentence test,". *Chin. J. Otorhinolaryngol* 1988;23:97–99.
- Huang TS, Wang NM, Liu SY. "Nucleus 22-channel cochlear mini-system implantations in Mandarin-speaking patients,". *Am. J. Otol* 1996;17:46–52. [PubMed: 8694134]
- Keppel, G. *Design and Analysis: A Researcher's Handbook*. 3rd ed.. Prentice-Hall; Englewood Cliffs, NJ: 1991.
- Liang ZA. "Tonal discrimination of Mandarin Chinese,". *Acta Physiol. Sinica* 1963;26:85–91.
- Lehiste, I. *Suprasegmentals*. MIT Press; Cambridge, MA: 1970.
- Loizou PC, Dorman M, Tu Z. "On the number of channels needed to understand speech,". *J. Acoust. Soc. Am* 1999;106:2097–2103. [PubMed: 10530032]
- Pfingst BE, Franck KH, Xu L, Bauer EM, Zwolan TA. "Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses,". *J. Assoc. Res. Otolaryngol* 2001;2:87–103. [PubMed: 11550528]
- Rosen S. "Temporal information in speech: Acoustic, auditory and linguistic aspects,". *Philos. Trans. R. Soc. London* 1992;336:367–373. [PubMed: 1354376]Ser. B
- Schouten JF, Ritsma RJ, Cardoz BL. "Pitch of the residue,". *J. Acoust. Soc. Am* 1962;34:1418–1424.
- Schroeder MR. "Reference signal for signal quality studies,". *J. Acoust. Soc. Am* 1968;44:1735–1736.
- Seligman P, McDermott H. "Architecture of the Spectra 22 speech processor,". *Ann. Otol. Rhinol. Laryngol* 1995;104(Suppl 166):139–141.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. "Speech recognition with primarily temporal cues,". *Science* 1995;270:303–304. [PubMed: 7569981]
- Shannon RV, Zeng FG, Wygonski J. "Speech recognition with altered spectral distribution of envelope cues,". *J. Acoust. Soc. Am* 1998;104:2467–2476. [PubMed: 10491708]
- Skinner MW, Clark GM, Whitford LA. "Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 Channel Cochlear Implant System,". *Am. J. Otol* 1994;15(Suppl 2):15–27. [PubMed: 8572106]
- Studebaker GA. "A rationalized arcsine transform,". *J. Speech Hear. Res* 1985;28:455–462. [PubMed: 4046587]
- Sun JC, Skinner MW, Liu SY, Wang FNM, Huang TS, Lin T. "Optimization of speech processor fitting strategies for Chinese-speaking cochlear implantees,". *Laryngoscope* 1998;108:560–568. [PubMed: 9546270]
- ter Keurs M, Festen JM, Plomp R. "Effect of spectral envelope smearing on speech reception. I,". *J. Acoust. Soc. Am* 1992;91:2872–2880. [PubMed: 1629480]
- ter Keurs M, Festen JM, Plomp R. "Effect of spectral envelope smearing on speech reception. II,". *J. Acoust. Soc. Am* 1993;93:1547–1552. [PubMed: 8473608]
- Thornton AR, Raffin MJ. "Speech-discrimination scores modeled as a binomial variable,". *J. Speech Hear. Res* 1978;21:507. [PubMed: 713519]
- Tyler RS, Preece JP, Lansing CR, Gantz BJ. "Natural vowel perception by patients with the Ineraid cochlear implant,". *Audiology* 1992;31:228–239. [PubMed: 1444934]
- Van Tasell DJ, Soli SD, Kirby VM, Widin GP. "Speech waveform envelope cues for consonant recognition,". *J. Acoust. Soc. Am* 1987;82:1152–1161. [PubMed: 3680774]
- van Wieringen A, Wouters J. "Natural vowel and consonant recognition by Laura cochlear implantees,". *Ear Hear* 1999;20:89–103. [PubMed: 10229511]

- Wei CH, Zeng FG, Cao KL, Wang ZZ. "Rate discrimination and tone recognition in Chinese cochlear implant patients,". *Assoc. Res. Otolaryngol. Abstr* 2001;24:293–293.
- Wei WI, Wong R, Hui Y, Au DKK, Wong BYK, Ho WK, Tsang A, Kung P, Chung E. "Chinese tonal language rehabilitation following cochlear implantation in children,". *Acta OtoLaryngol* 2000;120:218–221. [PubMed: 11603776]
- Whalen DH, Xu Y. "Information for Mandarin tones in the amplitude contour and in brief segments,". *Phonetica* 1992;49:25–47. [PubMed: 1603839]
- Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. "Better speech recognition with cochlear implants,". *Nature (London)* 1991;352:236–238. [PubMed: 1857418]
- Wilson BS, Rebscher S, Zeng FG, Shannon RV, Loeb GE, Lawson DT. "Design for an inexpensive but effective cochlear implant,". *Otolaryngol.-Head Neck Surg* 1998;118:235–241. [PubMed: 9482558]
- Xu SA, Dowell RC, Clark GM. "Results for Chinese and English in a multichannel cochlear implant patient,". *Ann. Otol. Rhinol. Laryngol* 1987;96(Suppl 128):126–127.
- Zeng FG. "Cochlear implants in China,". *Audiology* 1995;34:61–75. [PubMed: 8561684]
- Zhang, H.; Zhao, KL.; Wang, ZZ. MACC: Chinese Minimal Auditory Capability Test. Beijing Union Hospital; Beijing, P. R. China: 1988.

FIG. 1.

Acoustic features of the speech and artificial signals. (A) Raw waveforms (top row) and spectrograms (bottom row) of the Chinese syllable /xu/ (pronounced "shoo") spoken by a female. Panels from left to right show tone patterns 1 through 4. The lexical meaning associated with tones 1 through 4 are "void," "slowly," "permit," and "sequence," respectively. The darkness of the spectrograms represents the energy associated with time and frequency. The fundamental frequency and the harmonics of the voiced part (/u/) show flat, rising, falling/rising, and falling patterns for tones 1 through 4, respectively. The temporal envelopes of the waveforms also differ from one tone pattern to another. The durations of the syllables are about 0.6 s with the voiced part averaged around 0.4 s. (B) Spectrograms of the higher-pitched set of the frequency-modulated (FM) sweeps synthesized to mimic the four tone patterns of Mandarin Chinese. The fundamental frequencies are listed in Table I. The durations are constant at 0.5 s.

FIG. 2.

Tone, consonant, vowel, and sentence recognition as a function of number of channels. The upper panels plot the distribution of the percent-correct scores across all subjects in a boxplot format in which the three horizontal lines represent the 25th, 50th, and 75th percentiles and the ends of the vertical bars represent the minimum and maximum. Panels from left to right are for tone, consonant, vowel, and sentence recognition, respectively. The dashed line at 25% for tone, consonant, and vowel recognition indicates the chance performance. For sentences, chance performance was 0%. The number of subjects tested is indicated in the lower right corner of each panel. The lower panels show the statistical significance of pairwise comparison of the mean percent correct associated with number of channels as revealed by the Tukey test. The light- and dark-gray squares represent the significance levels at $p < 0.05$ and $p < 0.01$, respectively. The empty squares represent comparisons not statistically significant.

FIG. 3.

Subjective judgments of the sound quality as a function of number of channels. The subjective judgments of each subject were normalized to his or her highest judgment score across all tests. The boxplot shows the distribution of the mean normalized quality judgments of all nine subjects. In the boxplot, the three horizontal lines represent the 25th, 50th, and 75th percentiles, and the ends of the vertical line show the minimum and the maximum of the distribution.

FIG. 4.

Relationship between the subjective judgments of sound quality and the tone, consonant, vowel, and sentence recognition scores. Each dot represents percent correct from one speech test (ordinate) and the normalized quality judgment (abscissa). The correlation coefficients (r) are shown in the lower right corners of all panels.

FIG. 5. Distribution of tone recognition scores as a function of the low-pass cutoff frequency. In the boxplot, the three horizontal lines represent the 25th, 50th, and 75th percentiles and the ends of the vertical line show the minimum and the maximum of the distribution across all nine subjects. The group means are connected by the solid lines. The dashed line represents the chance performance at 25%. The upper and lower traces represent data obtained with 12 channels and 1 channel, respectively, as indicated by the labels.

FIG. 6.

Representation of the number-of-channels-versus-LPFs matrix of tone recognition scores of three individual subjects. Each panel shows the mean percent correct for tones from one subject. (A) Subject 3. (B) Subject 4. (C) Subject 5. For each panel, the abscissa and the ordinate represent the number of channels and the LPFs, respectively. The percent correct for tones, which ranged between 25% to 100%, is represented by the diameter of the filled circles as indicated by the scale bar at the top.

FIG. 7.

Representation of the pooled results for the number-of-channels-versus-LPFs matrix of tone recognition scores. The data are plotted in the contour format in which the percent correct is represented by the gray scale as indicated by the scale bar at the top. The abscissa and ordinate are both in logarithmic scales. (A) Data represent the average across all nine subjects who participated in the tone recognition tests using speech materials in which the syllable duration was not equalized. (B) Data represent the average across all four subjects who participated in the tone recognition tests using speech materials that had equal syllable duration. In both (A) and (B), a trade-off between the number of channels and the LPFs is evident by the gradient of the tone recognition scores along the main diagonal line.

FIG. 8. Syllable durations of tone 1 through 4. Each symbol represents duration of one syllable spoken either by a male voice (open square) or by a female voice (filled circles). The rightmost column, labeled “equal,” plots the durations of the syllables that were selected for equal durations for tones 1 through 4.

FIG. 9.

Mean recognition scores of the four-pattern FM sweeps across all four subjects. The data are plotted in the same format as in one of the panels in Fig. 7 except that the contours are plotted in coarser steps than they are in Fig. 7. Panels (A), (B), and (C) show the percent correct for the lower-, higher-, and both lower- and higher-pitched FM sweeps, respectively. The fundamental frequencies of the FM sweeps are listed in Table I.

TABLE I.

The fundamental frequencies (Hz) of the low- and high-pitched sets of four patterns of FM sweeps. The four patterns are (1) flat and high; (2) rising; (3) falling and then rising; and (4) falling, thus mimicking the tone patterns of Mandarin Chinese. Listed in the table are the initial and final frequencies of the FM sweeps. For FM pattern 3, the frequency changes direction in the middle of the sweep and that frequency is also listed in the table.

Patterns	Initial		Middle		Final	
	Low	High	Low	High	Low	High
(1) -	150	300	...	...	150	300
(2) /	75	150	...	...	150	300
(3) V	125	250	89	178	125	250
(4) \	150	300	...	...	75	150