Development and validation of the Mandarin speech perception test

Qian-Jie Fu,[a)]Meimei Zhu, and Xiaosong Wang
Division of Communication and Auditory Neuroscience, House Ear Institute,
2100 West Third Street, Los Angeles, California 90057
qfu@hei.org, mzhu@hei.org, xwang@hei.org

Abstract: Currently there are few standardized speech testing materials for Mandarin-speaking cochlear implant (CI) listeners. In this study, Mandarin speech perception (MSP) sentence test materials were developed and validated in normal-hearing subjects listening to acoustic simulations of CI processing. Percent distribution of vowels, consonants, and tones within each MSP sentence list was similar to that observed across commonly used Chinese characters. There was no significant difference in sentence recognition across sentence lists. Given the phonetic balancing within lists and the validation with spectrally degraded speech, the present MSP test materials may be useful for assessing speech performance of Mandarin-speaking CI listeners.
© 2011 Acoustical Society of America

1. Introduction

In the clinic, speech testing is used to determine whether a patient might benefit from devices such as hearing aids or cochlear implants (CIs). Accurate assessment using validated, standardized test materials is critical for evaluating the efficacy of the device and for guiding any device adjustments and/or auditory rehabilitation. With adult patients, speech testing generally includes monosyllabic word recognition as well as recognition of words in sentences. Commonly used test materials for English-speaking CI listeners include the Hearing in Noise Test (HINT) sentences (Nilsson *et al.*, 1994) and the City University of New York sentences (Boothroyd *et al.*, 1985). However, these materials were not specifically developed to evaluate CI performances. CI users experience poorer spectral resolution than do normal-hearing (NH) listeners, which may result in differences in speech understanding across test materials. More recently, Spahr and Dorman (2004) developed the AzBio sentence lists, which were balanced according to NH performance in quiet, while listening to a five-channel acoustic CI simulation.

Given the increasing numbers of Mandarin-speaking CI patients, there is a great need to develop standardized sentence materials that are rigorously validated for speech testing. Compared to English sentence materials, it is more difficult to phonetically balance Mandarin sentence materials as the three components (vowels, consonants, and Chinese tones) must be carefully considered. In Mandarin Chinese, fundamental frequency (F0) cues are important for lexical tone recognition (Lin, 1988), and lexical tone recognition is important for sentence recognition (Fu *et al.*, 1998). There are four tonal patterns in Mandarin Chinese, which are characterized by F0 contours: Tone 1 (flat F0), Tone 2 (rising F0), Tone 3 (falling–rising F0), and Tone 4 (falling F0). The same syllable /ma/ can mean "mother," "linen," "horse," or "scold" for Tones 1, 2, 3, or 4, respectively. A fifth tone (neutral tone or tone 0) is also occasionally used in Mandarin Chinese. The syllable /ma/ for the neutral tone means a question particle. Current CI technology only conveys weak F0 cues, which are encoded by

---

amplitude modulations in temporal envelope. Mandarin-speaking CI users' tone recognition depends strongly on other cues that co-vary with F0, e.g., amplitude contour, periodicity, and duration (Fu and Zeng, 2000).

The Mandarin Hearing in Noise Test (MHINT) was a first attempt toward developing standardized speech test materials for Mandarin-speaking listeners (Wong *et al.*, 2007). However, there are two major limitations for MHINT sentences for use with hearing impaired (HI) people or CI listeners. First, MHINT sentences were not phonetically balanced (in terms of vowels, consonants, and tones) within and across test lists. Second, list similarity and test–retest variability was validated only in NH listeners. Fu *et al.*, (1998) found that for NH subjects listening to acoustic CI simulations, Tone 3 (falling–rising) and Tone 4 (falling) could be more easily recognized than Tone 1 (flat) or Tone 2 (rising). Given the importance of lexical tones to Mandarin Chinese speech understanding and tone perception differences between NH and CI listeners, it may be more appropriate to validate list similarity in light of CI processing and perception, i.e., with limited spectral and temporal cues. Because spectral and temporal processing differs greatly across individual CI patients, acoustic simulations of CI processing (with limited spectral and temporal cues) may be helpful in designing appropriate test materials for CI users. Previous studies have shown that CI listeners can effectively access only four to eight spectral channels (Shannon *et al.*, 2004), and that CI performance is generally similar to that of NH subjects listening to a four-channel acoustic CI simulation (Li *et al.*, 2011). In this study, phonetically balanced Mandarin sentence test materials were developed and validated with NH subjects listening to unprocessed speech or speech processed by an acoustic simulation of four-channel CI processing. Although the four-channel CI simulation may not fully replicate the experience of electric hearing, the simulation allows for evaluation of the test materials under conditions of reduced spectral resolution and basal shift (as experienced by real CI users). The simulation also reduces the well-known intersubject variability in the real CI case, typically due to patient-specific factors (e.g., the proximity of electrodes to healthy neurons, the amount of CI experience, etc.).

2. Methods

2.1 Development of phonetically balanced lists

The Mandarin speech perception (MSP) sentence materials consist of ten lists of ten sentences each. Each sentence includes seven monosyllabic words. In developing the MSP materials, the first criterion was that the sentences should all be familiar and widely used in daily life. The second criterion was that each of the sentence lists should be phonetically balanced. The targeted number of vowels, consonants, and tones within each list was first computed according to the statistical distribution across 3500 commonly used Mandarin Chinese words (Tang, 1995). Due to the limited number of words (70) in each list, some variation of the number of vowels, consonants, and tones was allowed for each list; the number of targeted vowels and consonants within each list was allowed to vary by ± 1 and the number of targeted tones was allowed to vary by ± 2. For example, the rate of occurrence for the vowel /u/ is 8.43% across 3500 commonly used Chinese words, according to the statistical analyses (Tang, 1995). Given 70 words within a MSP sentence list, the target number of the occurrences for the vowel /u/ in each list is 6 (± 1). No unique word combinations of vowel, consonant, and tone were repeated within a list. The number of words repeated across lists was minimized (less than ten even for commonly used pronouns "he," "she," etc.). Disyllables were not repeated across lists. Figure 1 shows the distribution of vowels, consonants, and tones across 3500 commonly used Chinese characters (Tang, 1995) and the present MSP sentence lists. Table I shows an example MSP test list.

2.2 Recordings of sentence lists

After developing the phonetically balanced sentence lists, all sentences were clearly produced by a single female talker at a normal speaking rate. At the time of recording,
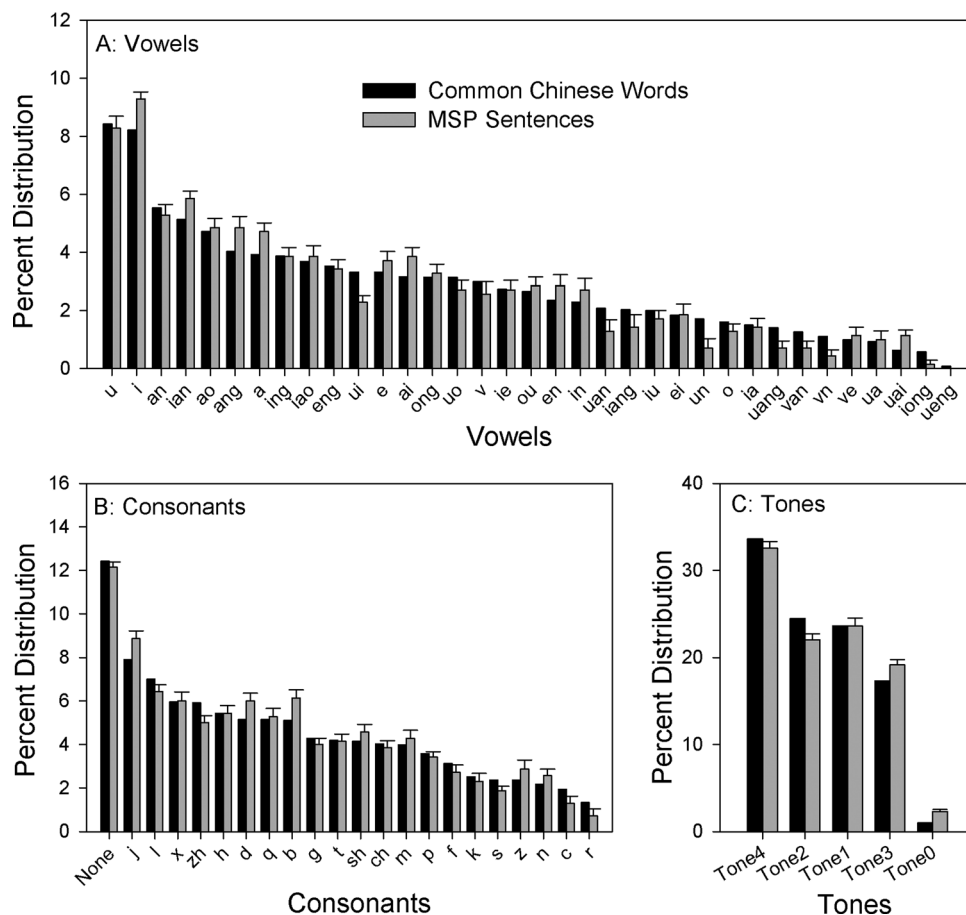
Fig. 1. The percent distribution of 35 vowels (A), 21 consonants (B), and five tones (C) across 3500 commonly used Chinese characters (black bars; data from Tang, 1995) and for the MSP sentence materials (gray bars). All the vowels and consonants were used according to the international standard *Scheme of the Chinese Phonetic Alphabet* (Yin and Felley, 1990; http://en.wikipedia.org/wiki/Pinyin). Note that "None" in (B) indicates the percent distribution of Chinese characters that have no initial consonants. Tones 0–4 represent neutral tone, flat tone, rising tone, falling–rising, and falling tone, respectively. Error bars show the standard error of the percent distribution across lists.

the talker had more than 10 yr of professional experience as a broadcaster in a radio station. Each sentence was recorded several times and the most clearly pronounced sentence was included in the test materials used for the validation study. For the recorded test materials, the mean sentence duration was 1974 ± 129 ms, the mean speaking rate was 3.55 ± 0.08 words/s, and the mean F0 was 223 ± 15 Hz. The audio recording of all sentence material can be downloaded and/or played at the following web site: http://www.tigerspeech.com/msp/msp.html.

2.3 Subjects

Twelve NH subjects (four males and eight females) participated in the validation of four-channel vocoded speech and eight NH subjects (four males and four females) participated in the validation of unprocessed speech. Subjects were native speakers of Mandarin Chinese and were between the ages of 20 and 48 yr old. All had thresholds better than 20 dB hearing level at audiometric frequencies from 0.25 to 8 kHz. All subjects were paid for their participation, and all provided informed consent in accordance with the local Institutional Review Board.

Table 1. Example of sentence test list.

#	Chinese character	Chinese pinyin	English translation
1	今天的阳光真好	jīn tiān de yáng guāng zhēn hǎo	It's a nice sunny day.
2	节假日不用门票	jié jiǎ rì bù yòng mén piào	No ticket is needed during holiday.
3	晚上一块去跳舞	wǎn shàng yī kuài qù tiào wǔ	Let's go dancing together tonight.
4	对面有两所高中	duì miàn yǒu liǎng suǒ gāo zhōng	There are two high schools across street.
5	这些衣服洗过吗	zhè xiē yī fú xǐ guò ma	Have these clothes been washed yet?
6	**北京近来很寒冷**	běi jīng jìn lái hěn hán lěng	It's very cold in Beijing recently.
7	他家每年放鞭炮	tā jiā měi nián fàng biān pào	He touches off the firecracker every year.
8	外孙出生在农村	wài sūn chū shēng zài nóng cūn	Grandson was born in rural areas.
9	星期二别打篮球	xīng qī èr bié dǎ lán qiú	Don't play basketball on Tuesday.
10	短裙长度正合适	duǎn qún cháng dù zhèng hé shì	The length of short skirt is appropriate

2.4 Signal processing

NH subjects were tested while listening to unprocessed speech or to a four-channel, sine-wave vocoded acoustic simulation of CI speech processing. A sine-wave vocoder was used instead of a noise-band vocoder because our recent studies suggest that sine-wave vocoders better correspond to CI performance for pitch related tasks, such as voice gender recognition (Fu *et al.*, 2005). For vocoded speech, the input acoustic signal was band-pass filtered into four frequency bands using fourth-order Butterworth filters. The cutoff frequencies of the analysis bands were 200, 591, 1426, 3205, and 7000 Hz, respectively. The amplitude envelope was extracted from each band by half-wave rectification and low-pass filtering (fourth-order Butterworth) with a 160 Hz cutoff frequency. The extracted envelope from each band was used to modulate sine-wave carriers whose center frequencies were the arithmetic center frequencies of the analysis bands. Finally, the modulated carriers were summed and normalized to have the same long-term root-mean-square as the input speech signal.

2.5 Procedures

Stimuli were presented in a sound field at 65 dBA via a single loudspeaker; subjects were seated directly facing the loudspeaker at a 1 m distance. Prior to formal testing, NH subjects listened to alternate speech materials (e.g., the MHINT sentences) processed by the four-channel CI simulation to minimize procedural learning (e.g., familiarization with the speech processing, the test procedures, environment, etc.). During testing, a sentence list was randomly selected and sentences were randomly selected from within the list (without replacement) and presented to the subject, who repeated the sentence as accurately as possible. Subjects were instructed to guess if they were not sure, but were cautioned not to provide the same response for each stimulus. The experimenter calculated the percent of words correctly identified in sentences. All words in the MSP materials were scored, resulting in a total of 70 words for each list. No training or trial-by-trial feedback was provided during testing. All lists were tested with each subject. The test order of the sentence lists was randomized and counterbalanced across subjects.

3. Results

NH subjects scored 100% correct with the original, unprocessed, MSP sentences. Figure 2 shows mean word-in-sentence recognition scores as a function of the MSP list number for NH subjects listening to a four-channel CI simulation. Mean word-in-sentence recognition across lists and subjects was 90.9% correct (range: 88.7%–93.1% correct); the mean standard error was 2.00% (range: 1.59%–2.35%). A one-way repeated-measures analysis of variance, with the test list as a treatment factor showed no significant effect for the test list [$F(9,119) = 1.756$, $p = 0.086$]. The standard
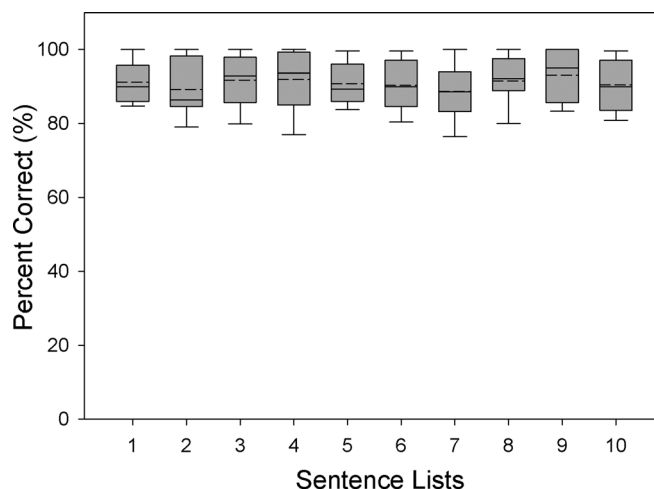
Fig. 2. Sentence recognition scores as a function of MSP sentence list. Twelve NH subjects were tested while listening to a four-channel CI simulation. The upper edge of the box indicates the 95th percentile of the data set, and the low edge indicates the 5th percentile. The solid lines in the box indicate the median recognition scores and the dashed lines indicate the mean recognition scores.

deviation across ten test lists ranged from 1.35% to 5.23% for individual subjects, with a mean of 3.30%.

4. Discussion

The MSP sentence materials were phonetically balanced and validated in NH subjects listening to unprocessed speech and a four-channel acoustic CI simulation. Although mean performance was approximately ten points poorer with the simulation than with unprocessed speech, there was no significant difference across lists in either processing condition. As such, the MSP sentences meet four important criteria for development of speech testing materials, namely, familiarity, homogeneity, phonetic balancing, and list similarity (Tsai *et al.*, 2009). As there are no standard testing materials with which to evaluate speech recognition performance in Mandarin-speaking CI users, the MSP sentences offer several clinical advantages.

First, the MSP materials include phonetically balanced sentence lists. This is the first Mandarin Chinese sentence database to include phonetically balanced materials, whether for testing NH, HI, or CI listeners. Phonetic balancing helps to ensure that sentence recognition testing represents listeners' speech understanding, given the distribution of vowels, consonants, and tones according to common Chinese words.

Second, the MSP materials were validated using NH subjects listening to unprocessed speech, as well as to a four-channel CI simulation. Many standard test materials (e.g., HINT sentences, MHINT sentences) have been validated using NH subjects only listening to unprocessed speech. This seems reasonable for comparing HI or CI performance to NH norms. However, CI norms may be quite different, as suggested previously in the simulation study by Fu *et al.*, (1998). If CI users are able to access only limited amounts of spectral and/or temporal cues, certain speech features (e.g., vowel formants, consonant articulations, tone directions) may be differently weighted. It is useful to see whether list similarity is affected by the availability of these cues. Most likely, for NH subjects listening to unprocessed speech, phonetic balancing will result in similar performance across lists. In the present study, mean performance with unprocessed speech was 100% correct for each list. When limited spectral and/or temporal cues are available (as in the present four-channel CI simulation), speech recognition performance may differ across lists, despite the phonetic balancing. In such a case, lists may be rebalanced to produce similar performance across lists. In the present

study, mean performance with the four-channel simulation was similar across lists. Thus, the MSP test lists are both phonetically and perceptually balanced whether with unprocessed speech or with a simulation of hearing impairment.

Third, the criteria for the MSP sentences may allow for a more valid assessment of speech perception performance. As described previously, no unique word combinations of vowel, consonant, and tone were repeated within a list, no disyllables were repeated across lists, and common words were minimally repeated across lists. As such, all words in each list could be considered to be "keywords," resulting in a total of 70 keywords per list. For MHINT sentences (Wong *et al.*, 2007), most of keywords are disyllables. There are only 90 keywords per list, even though there are 20 sentences per list and 10 words per sentence. From this perspective, the MSP allows for quicker evaluation because fewer words are needed. The fewer number of words and general ease of difficulty in the MSP materials may also allow for testing with children and noise.

The validation of speech materials is generally linked to the specific recording. However, different recordings of commonly used test materials [e.g., IEEE sentences (Rothauser *et al.*, 1969)] have been used in different studies without any validation of the recordings. Speech recordings may differ greatly in terms of speaking style, especially speaking rate. Recently, Li *et al.*, (2011) found that different speaking styles (such as speaking rate, whispering) produced by the same talker significantly affected recognition of easy sentences by Mandarin-speaking CI subjects. This suggests that having a common recording reference (as in the present MSP materials) will facilitate data comparison across studies. Alternate recordings/materials could be compared to this common reference. To facilitate the introduction of standardized sentence materials, the MSP sentences used in this study have been integrated within an open Window-based software platform. Both the testing platform and testing materials are freely available to researchers or clinicians (http://www.tigerspeech.com/msp/msp.html).

In this study, phonetically balanced Mandarin sentence test materials were developed and validated with NH subjects listening to unprocessed speech or speech processed by an acoustic simulation of four-channel CI processing. However, simulation studies may not perfectly predict real CI performance. Further validation of these sentence materials is needed with the HI populations and, specifically, with Mandarin-speaking Chinese CI users.

References and links

Boothroyd, A., Hanin, L., and Hnath, T. (**1985**). "A sentence test of speech perception: Reliability, set equivalence, and short term learning," Internal Report No. RCI 10, New York: City University of New York, New York.

Fu, Q.-J., Chinchilla, S., Nogaki, G., and Galvin, J. J., III (**2005**). "Voice gender discrimination: the role of periodicity and spectral profile," J. Acoust. Soc. Am. **118**, 1711–1718.

Fu, Q.-J., and Zeng, F.-G. (**2000**). "Effects of envelope cues on Mandarin Chinese tone recognition," Asia-Pacific J. Speech Lang. Hear. **5**, 45–57.

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

Li, Y., Zhang, G., Kang, H.-Y., Liu, S., Han, D. and Fu, Q.-J. (**2011**). "Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users," J. Acoust. Soc. Am. **129**(5), EL242–EL247.

Lin, M.-C. (**1988**). "The acoustic characteristics and perceptual cues of tones in Standard Chinese," Chinese Yuwen **204**, 182–193.

Nilsson, M. Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Rothauser, E. H., Chapman, N. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "IEEE recommended practice for speech quality measurements." IEEE Trans. Audio Electroacoust. **17**, 227–246.

Shannon, R. V., Fu, Q. J., and Galvin J. J. III (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Otolaryngol. Suppl. **552**, 50–54.

Spahr, A. J., and Dorman, M. F. (**2004**). "Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices," Arch Otolaryngol Head Neck Surg. **130**, 624–628.

Tang, Y.-H. (**1995**). "Statistical analysis of Mandarin Chinese," J. Chengde Teachers' Coll. Nat. **1995**, 66–76

Tsai, K. S., Tseng, L. H., Wu, C. J., and Young, S. T. (**2009**). "Development of a Mandarin monosyllable recognition test," Ear Hear. **30**, 90–99

Wong, L. L., Soli, S. D., Liu, S., Han, N., and Huang, M. W. (**2007**). "Development of the Mandarin hearing in noise test (MHINT)," Ear Hear. **28**, 70S–74S.

Yin, B., and Felley, M. (**1990**). *Chinese Romanization: Pronunciation and Orthography* (Sinolingua, Beijing).