



Published in final edited form as:

Acta Otolaryngol. 2012 August ; 132(8): 855–861. doi:10.3109/00016489.2011.653668.

Development and validation of the Mandarin disyllable recognition test

Meimei Zhu^{1,2}, Xiaosong Wang¹, and Qian-Jie Fu¹

¹Division of Communication and Auditory Neuroscience, House Research Institute, Los Angeles, USA

²Department of Otorhinolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai, China

Abstract

Conclusion—Given the phonetic balancing across lists and the validation with spectrally degraded speech, the present Mandarin disyllable recognition test (DRT) materials may be useful for assessing speech performance of Mandarin-speaking cochlear implant (CI) users. If combined with the previously developed sentence materials, these materials would help to establish standardized speech perception tests for Mandarin-speaking hearing-impaired (HI) and CI patients.

Objectives—To develop standardized Mandarin DRT materials that can be used to evaluate the speech performance of Mandarin-speaking HI and CI patients, and to establish standardized Mandarin speech perception test materials that include both disyllables and sentences.

Methods—Ten phonetically balanced Mandarin DRT lists were developed. The DRT materials were validated in 8 normal hearing (NH) subjects listening to unprocessed speech and in 10 NH subjects listening to a 4-channel, sine-wave vocoded acoustic simulation of CI speech processing. Performance with the DRT materials was compared to that with Mandarin sentence materials previously developed by our group.

Results—The distribution of vowels, consonants, and tones within each DRT list was similar to that observed across commonly used Chinese characters. There was no significant difference in disyllable word recognition across lists in both unprocessed and four-channel vocoded speech. There was a significant correlation between disyllable and sentence recognition performance.

Keywords

Cochlear implant; speech perception; phonetic balancing

Introduction

For hearing aid (HA) and cochlear implant (CI) users, it is important to accurately assess the efficacy of the hearing device using validated, standardized test materials. Standardized test

materials are also important for guiding device adjustments and/or auditory rehabilitation. Open-set word recognition is thought to accurately reflect individuals' speech perception capabilities, and is widely used to assess auditory communication in both clinical and research settings. Currently, there are several standard word recognition tests for English. Commonly used open-set word recognition test materials for adults include: phonetic balance 50 list (PB-50), Central Institute for the Deaf W-22 (CID W-22), and Northwestern University Auditory Tests Number 4 (NU-4) and 6 (NU-6). For children, tests include the PBK-50, lexical neighborhood test (LNT), and the multisyllable lexical neighborhood test (MLNT; featuring lexically easy and hard words).

However, currently there are no such standardized Mandarin testing materials in China. Given that Mandarin is spoken by the largest population in the world, and the numbers of Mandarin-speaking CI patients have increased sharply in recent years, it is urgent to develop standardized word recognition materials that are rigorously validated for Mandarin speech testing. Three criteria should be satisfied in the development of speech recognition testing material: familiarity, phonic balance, and equivalence [1]. Compared with English word recognition testing materials, it is more difficult to phonetically balance Mandarin word materials, as three dimensions – vowels, consonants, and Chinese tones – must be considered. The biggest difference between Mandarin and English is that the former is a tonal, Sino-Tibetan language, while the latter is a non-tonal, Indo-European language. 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 produced with different tones can have vastly different meanings [2]. For example, the same syllable /ba/ means 'eight,' 'pull,' 'target,' or 'dad' when produced by Tones 1, 2, 3, or 4, respectively. A fifth tone (neutral tone or Tone 0) is occasionally used in Mandarin Chinese. The syllable /ba/ with the neutral tone is often used as a mood particle at the end of a sentence. In Mandarin Chinese, fundamental frequency (F₀) cues are important for lexical tone recognition [3], and lexical tone recognition is also critical for recognition of Mandarin words. However, only weak F₀ cues, encoded by amplitude modulations in the temporal envelope, are conveyed by the current CI devices. Along with F₀, other cues such as amplitude contour, periodicity information, and duration also contribute to Mandarin-speaking CI users' tone recognition. Tones 3 and 4 are more easily recognized than Tones 1 and 2; thus, it is important to balance Chinese tones when developing Mandarin word test materials [4]. Chinese vowels are also more difficult to balance phonetically across lists than English since Mandarin has many more vowel phonemes (35) in common usage than does English (20).

Several Chinese word recognition tests have been developed. Yuen et al. [5] developed an open-set speech recognition test to assess pediatric Cantonese-speaking HA or CI users' speech recognition performance. Nissen et al. [6] also developed test materials that can be used to measure Cantonese word recognition. Mandarin trisyllabic words were developed to measure speech reception thresholds (SRTs) in quiet [7]. Several studies developed monosyllable word test materials that were validated using NH listeners [8–10]. As disyllables are widely used in daily life, it is important to develop standardized disyllable testing materials alongside these previous test materials.

It is more difficult to phonetically balance disyllables than monosyllables across test lists. While several groups have developed the disyllable test, some did the phonetic balance [11,12]. However, these previous studies used a probability distribution based on the how frequently the phoneme appeared in the daily life. For example, ('ta,' means 'he/she'), ('wo,' means 'I'), and ('de,' indicates possession) do not contribute much to speech understanding but have a high probability of occurrence. Such a distribution may not reflect patients' true speech recognition capabilities. In these previous test materials, many words were repeated within or across lists, which could create response bias due to repetition or memory effects. Also, none of the previous test materials have been evaluated for list equivalency in terms of degraded speech resolution, which is vital for standardizing such test materials for use in hearing impaired (HI) patients or CI users.

Recently, we developed phonetically balanced Mandarin sentence test materials and validated the materials with NH subjects listening to unprocessed speech and speech processed by an acoustic simulation of four-channel CI processing. No significant difference in sentence recognition for both unprocessed and processed speech was found across sentence lists [13]. Given the limited spectral and temporal resolution of the CI and given the variability in CI patient performance, it is useful to validate speech testing materials in NH listeners (whose performance would be presumably less variable) using acoustic stimulations of CI processing. The four-channel CI simulation allowed for evaluation of the test materials under similar signal processing conditions as experienced by real CI users. Only after the phonetic balancing within lists and the validation with spectrally degraded speech, the testing materials for both sentence and disyllable may be useful for assessing speech performance of Mandarin-speaking CI listeners. We also tested open-set disyllable and sentence recognition in 37 Mandarin pediatric CI users using some other test materials developed by others and found that while both pre- and post-lingually deafened pediatric CI patients were able to complete these somewhat difficult tests, word and sentence recognition performance was significantly different between the two subject groups [14]. As so far the Mandarin test materials were developed based on different standards and some of them still did not meet the basic speech recognition testing material development criteria, there was no comparability for the test results derived from different test materials. It is important to develop similar word lists with the same criteria as the sentences [13], so that these test materials can be used as the standard test materials to verify the word and sentence recognition performance in CI users.

The purpose of this study was to develop standardized Mandarin disyllable test materials that can be used to evaluate the speech performance of HI people and CI users. Combined with the previously developed sentence test materials [13], the present disyllable test materials will help to establish standardized Mandarin speech test materials. As in the previous study [13], disyllable materials were balanced across test lists and were validated using NH subjects listening to unprocessed speech or speech processed by a four-channel CI simulation.

Materials and methods

Development of phonetically balanced lists

The Mandarin disyllable recognition test (DRT) materials consist of 10 lists of 35 disyllables each. In developing the DRT materials, the first criterion was that the words should all be familiar and widely used in daily life. The second criterion was that the materials should include a representative distribution of nouns, verbs, adjectives, and adverbs found in everyday speech. The third criterion was that the disyllable 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 [15]. Due to the limited number of syllables (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 consonant /j/ is 7.91% across 3500 commonly used Chinese words [15]. Given 70 syllables within a DRT list, the target number of occurrences of /j/ is 5 (± 1) within each list. No unique word combinations of vowel, consonant, and tone were repeated within a list, and disyllables were not repeated across lists. The number of words (Mandarin characters) repeated across lists was less than five, and the number of Pinyin repeated across lists was less than six. Figure 1 shows the distribution of vowels, consonants, and tones across 3500 commonly used Chinese characters [15] and the present DRT lists. Table I shows an example of a DRT test list.

Recordings of disyllable lists

All test materials were clearly produced by a single female talker at a normal speaking rate. The talker had more than 10 years of professional experience as a broadcaster in a radio station. Each disyllable was recorded several times and the most clearly produced disyllable was included in the test materials for the validation study. For the recorded test materials, the mean disyllable duration was 932 ± 81 ms, the mean speaking rate was 2.16 ± 0.19 syllables per second, and the mean F0 was 208 ± 22 Hz. Audio recordings of all DRT materials can be found at: <http://www.tigerspeech.com/msp/dsp.html>.

Subjects

Ten NH subjects (three males and seven females) were tested while listening to four-channel vocoded speech, and eight of them (three males and five females) were tested while listening to unprocessed speech. Six of them also participated in our previous sentence test. Subjects were native speakers of Mandarin Chinese and were between the ages of 27 and 48 years. All subjects had pure tone thresholds < 15 dB HL for audiometric frequencies between 250 and 4000 Hz. All subjects were paid for their participation and all provided informed consent before testing was begun, in accordance with the local Institutional Review Board.

Signal processing

NH subjects were tested while listening to unprocessed speech or to speech processed by a four channel, sine-wave vocoded simulation of a CI. A sine-wave vocoder was used instead

of noise-band vocoder because our recent studies suggest that NH performance with sine-wave simulations better corresponded to real CI performance for pitch-related tasks such as voice gender recognition [16]. For vocoded speech, the input acoustic signal was band-pass filtered into four frequency bands using fourth-order Butterworth filters; the cut-off frequencies of the analysis bands were 200, 591, 1426, 3205, and 7000 Hz. The amplitude envelope was extracted from each band by half-wave rectification and low-pass filtering (fourth-order Butterworth with a 160 Hz envelope cut-off frequency). The extracted envelope from each band was then 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 (RMS) as the input speech signal.

Procedures

Stimuli were presented in the sound field at 65 dBA via a single loudspeaker; subjects were seated directly facing the loudspeaker at a 1 m distance. Before formal testing, NH subjects listened to alternate speech materials (e.g. the previously developed Mandarin Speech Perception, MSP, sentences) processed by the four-channel CI simulation to minimize procedural learning and to familiarize subjects with the speech processing, test procedures, test environment, etc. During testing, a test list was randomly selected, and disyllables were randomly selected from within the list (without replacement) and presented to the subject, who repeated the disyllable 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 percentage of syllables (or monosyllabic words) correctly identified in disyllabic words. All words in the DRT materials were scored, resulting in a total of 70 monosyllabic 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 disyllable lists was randomized and counterbalanced across subjects.

Results

Figure 2 shows the mean score of each DRT list for NH subjects listening to four-channel CI simulation and the mean score across lists and subjects for both processed and unprocessed speech. NH subjects scored 100% correct with the original, unprocessed DRT lists. Mean word recognition with the four-channel CI simulation (across lists and subjects) was 78.26% correct (range 76.43–79.29% correct). A one-way repeated-measures analysis of variance (RM ANOVA), with test list as treatment factor, showed no significant effect for test list [$F(9,81) = 0.781, p = 0.634$]. Six of the present subjects also participated in the previous validation study with Mandarin sentence materials [13]. Linear regression analyses were performed between mean sentence and mean disyllable performance with the four-channel CI simulation. Performance was highly correlated between sentences and disyllables ($r^2 = 0.897, p = 0.004$). A one-way RM ANOVA showed that sentence recognition was significantly better than disyllable recognition [$F(1,5) = 31.743, p = 0.002$].

Discussion

All the DRT materials were phonetically balanced according to the static standard of the statistical distribution across 3500 commonly used Mandarin Chinese words [15]. Compared with the dynamic standard of the guideline, the phonetic distribution for the static standard is more balanced. For the dynamic standard, /i;/e;/d/; and /sh/ account for much of the distribution of vowels and consonants, and /iong;/ueng/; and /uai/ are not represented at all. As the purpose of developing the test materials was to evaluate comprehension of all phonemes, the static standard was deemed as more appropriate for the design of the disyllable test materials. Also, the high incidence of /i;/e;/d/; and /sh/ are due to the frequent use of non-meaningful characters, e.g. [dè], which corresponds to the possessive form in English and [shì], which corresponds to am/is/are in English. These utterances occur frequently in spoken language, but contribute very little to understanding words and sentences. Consequently, performance with the present DRT materials represents listeners' everyday speech understanding, given the distribution of vowels, consonants, and tones according to common Chinese words.

In our study, DRT materials were validated in NH subjects listening to both unprocessed speech and a four-channel CI simulation. In most previous studies [6–10], word test materials were validated only in NH subjects listening to unprocessed speech. Given the importance of lexical tones to Mandarin Chinese, and given the poor F0 coding in CI processing, validating with NH subjects listening to unprocessed speech may not be appropriate. The limited spectral and/ or temporal cues available to CI users should be considered when validating testing materials, and lists should be balanced so as to produce equivalent performance across lists, whether with unprocessed speech or with a CI simulation. In this study, mean performance with unprocessed speech was 100% correct for each DRT list. While mean performance was poorer with CI simulation, there was no significant difference in performance across lists. This confirmed that the present DRT lists were appropriately balanced for both unprocessed speech and CI-mediated speech, making the materials appropriate as clinical assessment tools.

The DRT materials were designed and validated using the same criteria as the Mandarin speech perception sentence materials in Fu et al. [13], in the hope of establishing a series of standardized Mandarin speech perception test materials. Performance between disyllable and sentence test materials was highly correlated, as shown in Figure 3. Mean performance with unprocessed speech was 100% correct for both sets of materials when listening to unprocessed speech. When listening to the four-channel CI simulation, mean disyllable recognition was somewhat poorer than mean sentence recognition (78% vs 93% correct). Contextual cues most likely contributed to the better performance with the sentences. Previous studies [14,17,18] have also shown better open-set recognition of sentences than of words for post-lingually deafened adult or pediatric CI users. Furthermore, a significant correlation between the performance of DRT and that of sentence test materials was found, which means consistency and stability of these speech testing materials in terms of development of series of standardized Mandarin speech perception test materials. To facilitate introduction of standardized Mandarin speech perception test materials, the DRT materials used in this study have been integrated within an open Windows-based software

platform. Both the testing platform and testing materials are freely available to researchers or clinicians (<http://www.tigerspeech.com/msp/dsp.html>).

In this study, phonetically balanced Mandarin disyllable testing materials were developed and validated in NH subjects listening to unprocessed speech and speech processed by an acoustic simulation of four-channel CI processing. Combined with the previously developed sentence materials [13], these test materials help to establish standardized Mandarin speech perception test materials. These materials are currently being validated in Mandarin-speaking Chinese CI patients, with the ultimate goal of firmly establishing standardized assessment tools for Mandarin-speaking HI individuals, HA users, and CI patients. With good assessment tools, CI research and technology can improve signal processing to preserve cues important for Mandarin speech recognition.

Acknowledgments

The authors thank all the subjects who participated in this study. The authors also thank John J. Galvin III for editorial assistance. This work was partially supported by NIH grant DC004993.

References

1. Penrod, JP. Behavioral evaluation: peripheral hearing function: speech threshold and recognition/discrimination testing. In: Katz, J., editor. Handbook of clinical audiology. 4. Maryland: Williams & Wilkins; 1994. p. 147-64.
2. Wu JL, Yang HM. Speech perception of Mandarin Chinese speaking young children after cochlear implant use: effect of age at implantation. *Int J Pediatr Otorhinolaryngol.* 2003; 67:247–53. [PubMed: 12633924]
3. Lin MC. The acoustic characteristics and perceptual cues of tones in Standard Chinese. *Chinese Yuwen.* 1988; 204:182–93.
4. Fu QJ, Zeng FG, Shannon RV, Soli SD. Importance of tonal envelope cues in Chinese speech recognition. *J Acoust Soc Am.* 1998; 104:505–10. [PubMed: 9670541]
5. Yuen KC, Ng IH, Luk BP, Chan SK, Chan SC, Kwok IC, et al. The development of Cantonese Lexical Neighborhood Test: a pilot study. *Int J Pediatr Otorhinolaryngol.* 2008; 72:1121–9. [PubMed: 18485493]
6. Nissen SL, Harris RW, Channell RW, Conklin B, Kim M, Wong L. The development of psychometrically equivalent Cantonese speech audiometry materials. *Int J Audiol.* 2011; 50:191–201. [PubMed: 21319936]
7. Nissen SL, Harris RW, Jennings LJ, Eggett DL, Buck H. Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. *Int J Audiol.* 2005; 44:391–9. [PubMed: 16136789]
8. Han D, Wang S, Zhang H, Chen J, Jiang W, Mannell R, et al. Development of Mandarin monosyllabic speech test materials in China. *Int J Audiol.* 2009; 48:300–11. [PubMed: 19842805]
9. Tsai KS, Tseng LH, Wu CJ, Young ST. Development of a mandarin monosyllable recognition test. *Ear Hear.* 2009; 30:90–9. [PubMed: 19125031]
10. Ji F, Xi X, Chen AT, Zhao WL, Zhang X, Ni YF, et al. Development of a Mandarin monosyllable test material with homogenous items (II): lists equivalence evaluation. *Acta Otolaryngol.* 2011; 131:1051–60. [PubMed: 21599549]
11. Wang S, Mannell R, Newall P, Zhang H, Han D. Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. *Int J Audiol.* 2007; 46:719–31. [PubMed: 18049961]
12. Nissen SL, Harris RW, Jennings LJ, Eggett DL, Buck H. Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *Int J Audiol.* 2005; 44:379–90. [PubMed: 16136788]

13. Fu QJ, Zhu M, Wang X. Development and validation of the Mandarin speech perception test. *J Acoust Soc Am*. 2011; 129:EL267–73. [PubMed: 21682363]
14. Zhu M, Fu QJ, Galvin JJ 3rd, Jiang Y, Xu J, Xu C, et al. Mandarin Chinese speech recognition by pediatric cochlear implant users. *Int J Pediatr Otorhinolaryngol*. 2011; 75:793–800. [PubMed: 21489643]
15. Tang YH. Statistical analysis of Mandarin Chinese patients with auditory neuropathy. *J Chengde Teachers' Coll Nationalities*. 1995; 1:66–76.
16. Fu QJ, Chinchilla S, Nogaki G, Galvin JJ 3rd. Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. *J Acoust Soc Am*. 2005; 118:1711–18. [PubMed: 16240829]
17. Gstoettner WK, Hamzavi J, Baumgartner WD. Speech discrimination scores of postlingually deaf adults implanted with the Combi 40 cochlear implant. *Acta Otolaryngol*. 1998; 118:640–5. [PubMed: 9840498]
18. Hamzavi J, Baumgartner WD, Pok SM, Franz P, Gstoettner W. Variables affecting speech perception in postlingually deaf adults following cochlear implantation. *Acta Otolaryngol*. 2003; 123:493–8. [PubMed: 12797584]
19. Yin, B., Felley, M. Chinese Romanization: pronunciation and orthography. Beijing: Sinolingua; 1990.

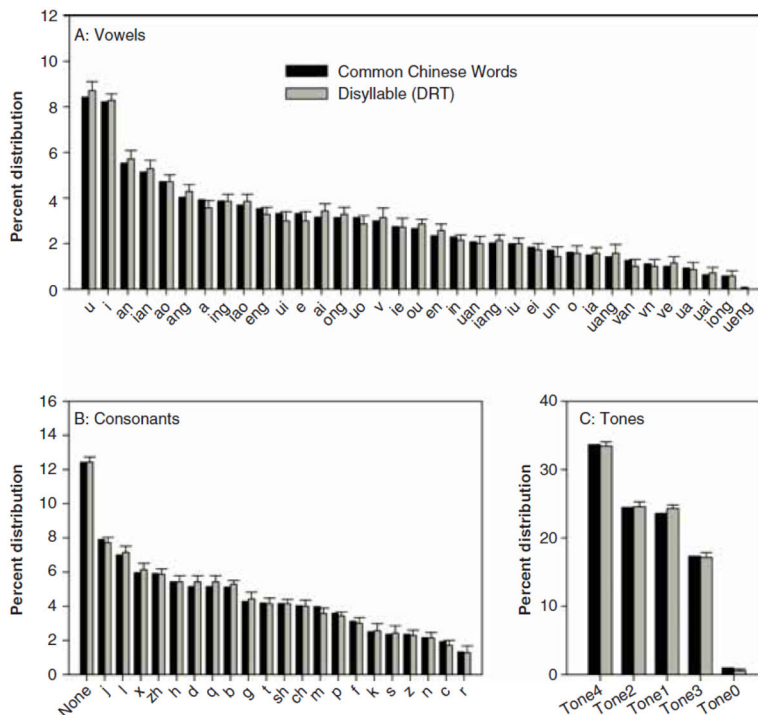


Figure 1. The percent distribution of 35 vowels (A), 21 consonants (B), and 5 tones (C) across 3500 commonly used Chinese characters (black bars; data from Tang [15]) and for the DRT materials (gray bars). All the vowels and consonants were used according to the international standard Scheme of the Chinese Phonetic Alphabet ([19]; <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, flat, rising, falling-rising, and falling tones, respectively. Error bars show the standard error of the percent distribution across lists.

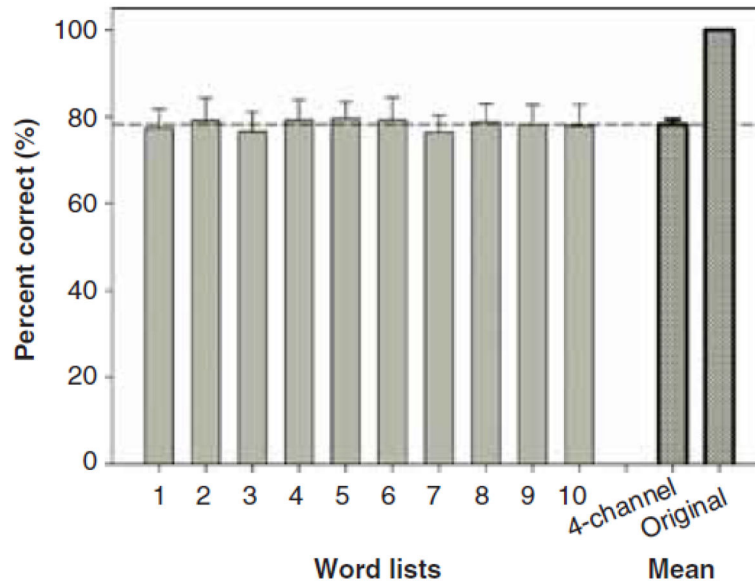


Figure 2. The mean score of each disyllable recognition test (DRT) list for normal-hearing (NH) subjects listening to fourchannel cochlear implant (CI) simulation and the mean score across lists and subjects for both processed and unprocessed speech. Error bars show the standard error (mean + SE).

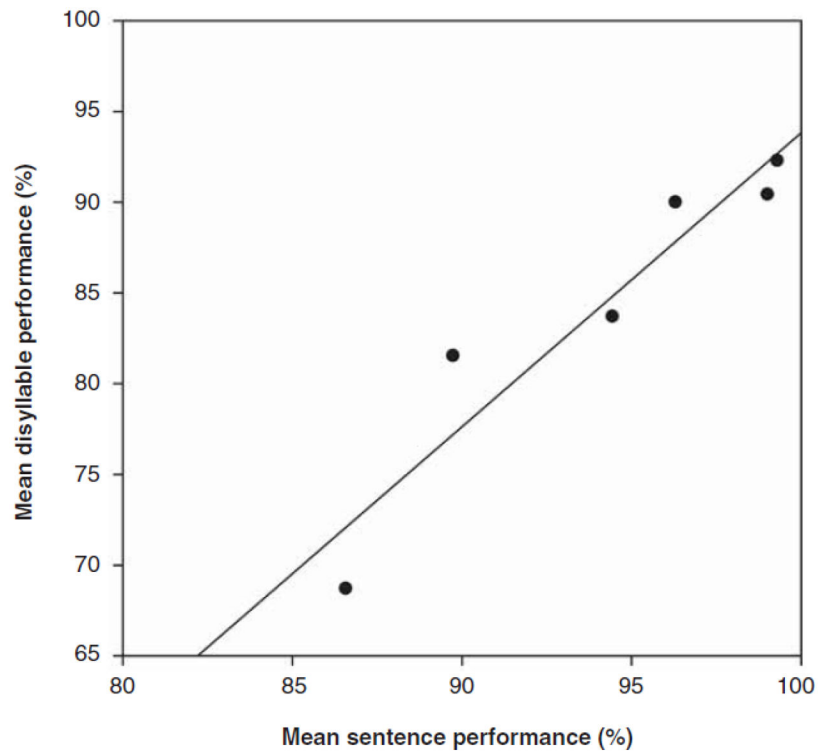


Figure 3. Mean disyllable recognition as a function of mean sentence recognition for six subjects who participated in the present study and in the previous Mandarin sentence validation study [13]. Performance while listening to the four-channel CI simulation is shown. The diagonal line shows the linear regression of the data.

Table 1

Disyllable testing list 5.

Number	Mandarin	Pinyin	English meaning
1	图片	túpiàn	Picture
2	问题	wèntí	Question
3	铅笔	qiānbǐ	Pencil
4	橡皮	xiāngpí	Eraser
5	窗户	chuānghù	Window
6	厕所	cèsuǒ	Bathroom
7	学历	xuélì	Educational background
8	运送	yùnsòng	Transport
9	鸡蛋	jīdàn	Egg
10	滚动	gǔndòng	Roll
11	冰箱	bīngxiāng	Refrigerator
12	萝卜	luóbo	Radish
13	增加	zēngjiā	Increase
14	轮船	lúnchuán	Ship
15	黄瓜	huángguā	Cucumber

Number	Mandarin	Pinyin	English meaning
16	杂物	záwù	Litter
17	夜晚	yèwǎn	Night
18	头发	tóufǎ	Hair
19	咳嗽	késòu	Cough
20	摔倒	shuāidǎo	Fall down
21	蜘蛛	zhīzhū	Spider
22	伙伴	huǒbàn	Partner
23	道路	dàolù	Road
24	民主	mínzhǔ	Democracy
25	去年	qùnián	Last year
26	需要	xūyào	Need
27	而且	érqiě	Furthermore
28	长度	chángdù	Length
29	掌握	zhǎngwò	Grasp
30	浇水	jiǎoshuǐ	Watering
31	工具	gōngjù	Tool

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Number	Mandarin	Pinyin	English meaning
32	敲门	qi omén	Knock on a door
33	流泪	liúlèi	Weep
34	安静	nǎng	Quite
35	牛奶	niú nǎi	Milk