Computers in Biology and Medicine 41 (2011) 131–138

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Development and evaluation of a computerized Mandarin speech test system in China

Wufang Wu [a,*], Hua Zhang [b], Jing Chen [b], Jianyong Chen [b], Changyan Lin [c]

[a] Capital Medical University, 100069, PR China
[b] Beijing Institute of Otolaryngology, 100730, PR China
[c] Beijing Anzhen Hospital of Capital Medical University, 100029, PR China

ARTICLE INFO

Article history:
Received 19 March 2010
Accepted 8 January 2011

Keywords:
Mandarin
Reliability
Validity
Sensitivity
Monosyllable
Disyllable
Sentence
Phonological balance
Digitally recorded
Computerized
Speech test

ABSTRACT

Objectives: This study reports the development and evaluation of a Computerized Mandarin Speech Test System (CMSTS). *Methods*: Taking into account the rules for developing speech materials and the unique linguistic characteristics of Mandarin, we designed and digitally recorded a set of materials comprised of seven lists of monosyllabic words, nine lists of disyllabic words, and fifteen lists of sentences with a high degree of subject familiarity. The CMSTS was developed with Visual Studio 2008, Access 2003 and DirectX 9. The system included five functions: listener management, a speech test, list management, data management, and system settings. We used the system to measure the speech recognition threshold (SRT) of 76 participants with normal hearing (age range: 20–28 years), and measured performance-intensity functions (PI) for all stimuli. *Results*: The SRT results were in accord with pure-tone results obtained by pure-tone audiometry. In a speech recognition score (SRS) test, changing the presentation level had the strongest effect on sentence recognition, followed by the presence of disyllabic words. Monosyllabic words were least affected by changes in presentation level. The slopes of the linear portion of the PI using the system were in accord with the findings of previous studies using audiometers and CDs with similar materials. *Conclusion*: The CMSTS has sufficient sensitivity, and can facilitate the wider use of speech audiometry in Chinese audiology clinics.

1. Introduction

Language is unique to humans, and is employed to express meaning, exchange ideas, and convey information. Speech is the most basic and important form of language, but diseases of the auditory system can affect the ability to understand spoken language and impede its basic functioning. In clinical studies and experimental research on hearing, it is important for physicians and researchers to understand the influence of hearing impairment on the level of speech recognition of individual patients. Speech audiometry is an intuitive and effective tool for assessing speech recognition. Traditional speech audiometry techniques use digital recordings on compact disc (CD) format in conjunction with an audiometer. Using these methods, the control of sound, item sequence, recording, and the analysis of test results are performed manually, resulting in a cumbersome and time-consuming testing process. In addition, the testing costs of traditional methods are high, and the results are easily affected by subjective factors related to both testers and participants. Moreover, the results of these testing methods are not typically conducive to statistical analysis. Recent developments in computer technology have enabled the development of digital intelligent speech audiometry systems. For example, Mackersie and Miniear [1] developed the Computer-Assisted Speech Perception Assessment Test (CASPA). The widely used Hearing in Noise Test (HINT) was designed primarily to measure recognition thresholds in the perception of spoken sentences. Versions of the HINT have been developed in many languages, including Swedish, Japanese, American English, Norwegian, and Korean [2–6].

Wong and Soli [7] developed a Cantonese version of the HINT, based on the English-language version. In addition, a Mandarin version was recently developed ([8]; MHINT), as well as a mainland Chinese version [9] and a Taiwanese version [10]. Because responses in the HINT are considered correct only when every word of a sentence is repeated correctly, this measure may ignore some phonetic information regarding key words. Even versions of the HINT using Chinese test materials involve a user interface in English. This constitutes an impediment to the wider use of the HINT in clinical work throughout China, especially in remote regions.

Chinese is a tonal language, with unique linguistic characteristics that differ markedly from other languages. In a clinical setting, the

* Corresponding author.
E-mail addresses: wuwenfang99718@yahoo.com.cn (W. Wu), a-zhang@263.net (H. Zhang), apple1470@sina.com (J. Chen), chenjianyong0506@sina.com (J. Chen), llbl@sina.com (C. Lin).

0010-4825/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2011.01.002

number of spoken words correctly recognized (the speech recognition score) as well as the recognition rate of critical information in given Chinese sentences are important in determining an accurate diagnosis. With this consideration in mind, we developed a computerized Mandarin speech test system, and evaluated the feasibility, reliability, and stability of the system.

Compared to HINT, this speech test system is able to examine the speech recognition threshold, speech recognition score, and speech recognition in noise. Moreover, this system includes monosyllabic words, disyllabic words, and sentence tests.

2. Methods

2.1. Participants

Seventy-six adults (41 females and 35 males) with normal hearing participated in this study. All participants spoke Mandarin at a functional level in their daily lives. Participants' ages ranged from 20 to 28 years, with a mean age of 23.9 years. All participants had pure-tone thresholds of less than 20 dB HL for octave frequencies between 250 and 8000 Hz. Medical histories revealed no otologic or hearing disorders, and only the ear with the best hearing threshold for each subject was tested, with 40 left ears and 36 right ears being assessed.

2.2. Test stimuli

We used Mandarin Speech Test Materials (MSTMs, [11]) as speech stimuli in this system. The content of the MSTMs is extremely general, and involves three types of stimuli: monosyllabic words, disyllabic words, and sentence stimuli. The monosyllabic test items consisted of one practice list of 10 words and seven equivalent lists of 50 words in each list. Disyllabic test items consisted of one practice list of 10 words and nine equivalent lists of 50 words each. Sentence test items consisted of one practice list having five sentences and 15 equivalent lists having 10 sentences (50 key words) each. The use of equivalent lists can reduce the effects of learning on the test, and the use of practice lists can reduce the initialization effect, in which participants' inexperience can lead to disproportionately low scores at the beginning of the test.

MSTMs were designed in accord with the basic rules for the design of speech materials, taking into account the unique linguistic characteristics of Mandarin. First, all the words were familiar to the participants, each chosen from well-established resources describing the most common words in the Chinese language, including The Frequency Dictionary of the Modern Chinese Language [12], and The Common Words of the Modern Chinese Language (National Language and Literature Committee, 1988). Second, all the words in each list were phonetically balanced [13]. Mandarin consists of 23 consonants, 38 finals, and four tones. The phonetic balances of these three components were all considered, and we ensured that the occurrence of consonants, finals, and tones was largely similar to that in speech in daily life. Third, sentence materials were consistent with spoken Chinese sentences, which represented whole samples of oral sentences commonly spoken in everyday life. Fourth, in accord with English sentence recognition tests such as the Central Institute for the Deaf (CID), Speech Perception in Noise (SPIN), and quick Speech in Noise (SIN), we recorded the recognition of key words in each sentence. Because participants could grasp the full meaning of the sentence if all key words were understood, the level of difficulty of the key words determined the difficulty in comprehending the test sentences as a whole.

The materials were recorded in a standard recording studio. The level of ambient noise was 25 dB (A), measured using a RION NL-11 sound-level meter. The materials were recorded using an Electro-Voice RE20 (USA) microphone connected to a LangXun digital audio station. The recorded materials were converted to CD format. The speaker was an experienced male Mandarin broadcaster approximately 30 years of age. The speaker was seated in an acoustically treated room and was asked to pronounce the words clearly and naturally, and to keep the intensity of speech sounds at a similar level. Finally, a 30-s calibration tone (1000-Hz pure tone) was inserted at the beginning of the recording [14]. A 4-s interval was inserted between the words and sentences. Items were then separated into single wave files on the CD. The RMS-level of the speech sounds in each file was adjusted to within ± 0.5 dB of the standard reference 1000 Hz tone.

2.3. Test system

Using Visual Studio 2008, Access 2003 and DirectX 9, we developed the computerized Mandarin speech test system. The system included five functions: listener management, speech test, list management, data management, and system settings. The architecture of the system is shown in Fig. 1.

2.3.1. Listener management

In this system, the tester is able to search, review, add, revise, and delete the listener's information. This information includes basic information as well as hearing functions that contain the results of pure tone air and bone conduction, deafness nature, hearing loss position, audiogram type, auditory brainstem response (ABR) threshold, distortion product otoacoustic emission (DPOAE), tympanogram type, and acoustic emission.

2.3.2. Speech test

The speech test included a measurement of speech recognition score (SRS) and speech recognition threshold (SRT), in an open set test. Using this system, testers can measure speech recognition in the presence of white or speech noise. Noise and speech can be delivered through separate channels (stereo) or mixed into a single channel (mono). The levels of both speech and noise can be adjusted independently. This system permits the randomization of the lists' order of presentation as well as randomization of the items within the lists to decrease learning effects.

Fig. 2 shows an example of a test screen as it appears to the tester during test administration. Following the presentation of a test item, the tester types the listener's response into the computer. The computer then displays both the test item and the listener's response. The tester selects the box corresponding to the repeated word using the mouse or by pressing the "tab" key. For each list, the system shows the ratio of words repeated correctly. Clicking the "next" button shows the next item, for selecting the test mode of manual control to allow the tester to select the most appropriate screen duration. If the automatic mode is selected, the system can automatically show the next item according to the time interval set by testers. If listeners require a rest, the "suspend" button can be clicked to pause the test, and the "continue" button resumes testing.

At the end of a list or series of tests, both the stimuli and corresponding subject responses for each test item are recorded onto the hard drive. In addition, the system can report and print the results in a graphic format (performance-intensity functions, as shown in Fig. 3). In addition, the results can be read into a "Microsoft Word" file for other analyses (Fig. 4).

When SRT is tested, the system can automatically increase or decrease the presentation levels according to the listener's