



Research Paper

Development of SC-10: A psychometrically equivalent Singapore Mandarin disyllabic word list for clinical speech audiometry use

Gary Jek Chong Lee ^{a,b,*}, Steven Lock Hey Lee ^c

^a Audiology, Ng Teng Fong General Hospital, Singapore

^b Temasek Skillsfuture Academy, Temasek Polytechnic, Singapore

^c Otolaryngology, Changi General Hospital, Singapore

Received 27 September 2019; received in revised form 23 January 2020; accepted 16 February 2020

Available online 23 July 2020

KEYWORDS

Speech audiometry;
Word recognition
score;
Disyllabic word list;
Mandarin;
Chinese;
Singapore;
SC-10

Abstract *Objective:* The aim of this study is to develop and evaluate a set of psychometrically equivalent disyllabic wordlist (SC-10) in Singapore Mandarin for clinical use.

Study design: A preliminary set of 1000 words were obtained from a dictionary of frequently used words by Singapore students. Ten native judges rate the familiarity level of each word. This is followed by a face-to-face public survey to rank the shortlisted word set from most to least familiar. The final 108 disyllabic words were recorded by a native female talker. 20 normal hearing subjects were used to obtain the percentage of correct word recognition at 24 intensity levels (−10 dB HL to 26 dB HL in 2 dB increment). Psychometric function slopes were calculated for each word. 100 words were eventually chosen and assigned into ten 10-word lists based on a psychometric balancing protocol. Minor digital adjustments were made to the intensity of each wordlist to improve their auditory homogeneity.

The developed SC-10 wordlists were validated by a separate group of 25 normal hearing subjects. Test-retest reliability was carried out on 20 out of 25 participants at the selected intensity levels (SRT-5, SRT, SRT+5).

Results: The calculated regression slopes in the psychometric functions for the ten lists are between 8.0 and 9.8%/dB. Single factor ANOVA analysis showed no significant difference in both the mean intensity required to obtain 50% recognition score ($f = 0.109$, $df = 9$, $p = 0.999$) and the slopes of the psychometric functions ($f = 0.078$, $df = 9$, $p = 0.999$) between the ten word lists. List validation on 25 normal hearing participants (PTA = 11.0 dB HL, $SD = 4.3$) showed a

* Corresponding author. Head, Audiology, Ng Teng Fong General Hospital, 1 Jurong East Street 21, 609609, Singapore.

E-mail address: Gary_LEE@nuhs.edu.sg (G.J.C. Lee).

Peer review under responsibility of Chinese Medical Association.



Production and Hosting by Elsevier on behalf of KeAi

mean speech recognition threshold (SRT) of 9.3 dB HL ($SD = 3.5$) and regression slope of 8.395%/dB. Quadratic regression analysis showed a positive correlation ($r^2 = 0.923$) between presentation level and word recognition score (WRS). The difference between PTA and SRT of each subject all fall within the clinically acceptable difference of 10 dB HL. Test-retest reliability, carried out on 20 subjects at three levels (SRT-5, SRT, and SRT+5 dB), showed no significance difference between word recognition score when the same participant is tested again at the same intensity level using a different wordlist.

Conclusion: All in all, it shows that the SC-10 speech materials are valid for clinical use for Mandarin speech audiometry in Singapore.

Copyright © 2020 Chinese Medical Association. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Speech audiometry uses speech stimuli to assess speech recognition as a means to evaluate hearing ability.¹ Besides assessing speech recognition, speech audiometry in quiet can be used to confirm pure tone audiometry thresholds,² to identify the site of lesion along the hearing pathway,³ to evaluate hearing aid benefits and to assess cochlear implant candidacy.⁴ Speech audiometry in quiet forms an integral part of diagnostic audiology test battery and is routinely conducted in audiology clinics in North America, Europe and Australia.^{5,6} However, speech audiometry is not routinely conducted in Singapore due to the lack of suitable speech audiometry materials for the diverse languages being spoken in Singapore.

Speech audiometry materials have been developed in various languages including Mandarin. Mandarin is spoken not only in People's Republic of China (where it is the official language), but also in countries with substantial Chinese populations, such as Malaysia and Singapore.⁷ Mandarin is a tonal language that contains 23 initial consonants, 38 finals and four tones with each of the four suprasegmental tones carries meaning.⁸ The 38 finals consist of 9 monophthongs, 13 diphthongs and triphthongs, with the remaining 16 being nasals.⁹ Mandarin speech audiometry materials have been developed in China since the 1950s. The history and development of Mandarin speech audiometry materials are well documented in literature.¹⁰ However, there has not been any Mandarin speech audiometry materials developed for the Singapore population to date.

Among the older generation in Singapore, a substantial percentage of its population is only literate in only one language; it is estimated that there are 472,200 individuals literate in only Mandarin.^{11,12} Given that the prevalence of hearing loss amongst the elderly in Singapore is high at 63.7%,¹³ appropriate speech audiometry materials need to be developed to support the hearing diagnosis and remediation of this population group.

Singapore Mandarin has been found to differ from Standard Mandarin in terms of pronunciation and tone.^{11,14,15} Ng¹⁵ found that the lip rounding required in producing /ü/ is often omitted in Singapore Mandarin. As a result, some words which can be clearly distinguished in Standard Mandarin pronunciation may not be easily told apart in Singapore Mandarin. Chen also observed that there

is a tendency to replace palatals with dentals in Singapore Mandarin, wherein the palatals "j, q, x" in Hanyu Pinyin are more often pronounced similarly to the dentals "z, c, s" instead. In addition, Ng¹⁵ found that the retroflex /sh/, which is constantly preserved when used as a consonant in Standard Mandarin, is uncommon in Singapore Mandarin. These differences have the potential to affect what a listener expects to hear and how his or her response is scored in speech audiometry, therefore motivates the necessity of a Singaporean Mandarin speech test material.

Other than pronunciation differences, there exist substantial lexical differences between Singapore Mandarin and Mandarin spoken in China. Mandarin is being referred to as Pǔtōnghuà (普通话) (Standard Mandarin) in mainland China and Huáyǔ (华语) (Singapore Mandarin) in Singapore. "乘客 chéng kè" (passenger) and "角落 jiǎo luò" (corner) in Standard Mandarin would likely be expressed differently in Singapore Mandarin as "搭客 dā kè" and "角头 jiǎo tóu" respectively.¹⁶ Chua¹⁶ noted that this could be attributed to the "borrowing" of vocabulary or transliterations from other dialects and languages spoken in the multi-racial society of Singapore. Word familiarity is an important consideration in designing speech audiometry materials,^{17,18} as studies have found this makes it easier for the words to be correctly identified.¹⁹

The unavailability of Singapore Mandarin Speech materials has resulted in no consistent approach towards speech audiometry across clinics in Singapore. With different clinics and clinicians adopting either Standard Mandarin or Malaysian Mandarin and presenting them over live voice or in the recorded accent of origin, it becomes difficult to compare outcomes across different clinics. This study aims to address the development of linguistically appropriate, valid and clinically efficient Singapore Mandarin Audiometric test materials for clinical use.

Disyllabic words were chosen for the development of the Singapore Mandarin speech audiometry materials as Mandarin words are seldom monosyllabic with over four-fifths of Mandarin words appearing in disyllabic form.²⁰ The word lists were formulated based on word familiarity rather than phonological balance. The importance of having a phonetically balanced word list for speech audiometry materials remains disputed, both in English and other languages.^{21–23} However, to maintain phonological balance, each wordlist has to be 50 words long to accurately represent the

occurrences of the various phonemes in everyday speech, which makes it less practical for clinical use. The authors have decided to construct, produce and validate ten psychometrically equivalent wordlists, each consisting of ten familiar Singapore Mandarin disyllabic words (SC-10 list) for use for the local population.

Material and methods

1 Words selection

The development of the wordlist was carried out at Temasek Polytechnic, a tertiary educational institution in the Republic of Singapore. The project scope and method was approved by the institution through the Head of Biomedical Informatics diploma unit. A preliminary set of 1000 commonly-used Mandarin disyllabic words were obtained from The Frequency Dictionary of Daily Chinese Words Encountered by Singapore Students.²⁴ A two-stage elimination process was used to shortlist the word list. First, ten native speakers of Singapore Mandarin were chosen as judges to rate each of the 1000 words accordingly to the familiarity level on a scale of 1–5 (1 = extremely familiar, 2 = very familiar, 3 = familiar, 4 = unfamiliar, 5 = rarely used). Words that were perceived to be inappropriate in meaning or possibly misunderstood as another homophonic word were also flagged as unsuitable. From the collated results, 700 words were eliminated if they rated (4) and (5) by at least two judges or flagged as unsuitable by any judge. With the remaining 300 words, a public face-to-face survey was administered at five community centres throughout the country. 25 randomly recruited public participants aged between 16 and 80 years of age were asked to rank the remaining word list from the most familiar to the least familiar. A final 108 Mandarin disyllabic words, representing the list of most commonly used Singaporean Mandarin words were shortlisted for recording.

2 Digital recording and production

One female speaker was selected from a group of four native Singapore-accented Mandarin speakers to record the word lists. The speakers were recruited from an advertisement posted via the toastmasters club at Temasek Polytechnic. An audition was carried out with the speakers' recorded voices evaluated by three judges. The speaker with the highest score based on voice quality, standard accent and pronunciation was selected as speaker for the recording of the word list.

The final speech material recording was conducted in a double-walled sound proof recording studio at the communications department at Temasek Polytechnic. A MD421U Sennheiser microphone, positioned about 6 cm from the talker at a 0° azimuth and covered by a 20 cm windscreen, was utilized for all recordings. The signal was digitized using a Yamaha audio capture card with a 44.1 kHz sampling rate with 24-bit quantization. During the recording sessions, the speaker was asked to pronounce each disyllabic word four times. The first and last repetition of each word was excluded to avoid any possible list effects, with the best production of the word selected. The

intensity of each disyllabic word was then digitally adjusted to yield the same average Root Mean Square (RMS) power as that of a 1 kHz calibration tone to equate test word threshold audibility.²⁵ This entailed the use of the adobe soundroom software. Each recorded word was then saved in a discrete.wav file and used for subsequent testing.

3 Formulation of the SC-10 wordlist

Twenty native Singaporean Mandarin users (9 male and 11 female) were recruited for a first audiological evaluation of all words, aimed at constructing the word lists. The subjects' ages ranged from 18 to 25 years ($M = 21.2$ years). All of the participants in this study had no reported otologic conditions and exhibited PTA thresholds ≤ 15 dB HL at octave frequencies from 250 to 8000 Hz. The mean arithmetic average of PTA thresholds at 500, 1000, and 2000 Hz for the 20 subjects was 10.2 dB HL (Table 1).

The tests were performed by three tertiary students from Temasek Polytechnic, who have undergone supervised training in basic audiometry. Subjects were tested in a double-walled sound booth at the Biomedical Informatics Diploma Unit, Temasek Polytechnic which met ANSI S3.1 standards. The testing of the word list as well as pure-tone audiometry was conducted using a Siemens SD28 diagnostic audiometer. The audiometer had undergone acoustic calibration by an external vendor one month prior to the study to ensure that its accuracy is within the limits set by the American Standard Specification for Audiometers, S3.6-1969. A biological calibration was carried out at the start of each day before tests to ensure that the audiometer to be functioning properly. As the results of the biological calibration were always within the 5 dB of the testers own known threshold on every occasion, no changes in calibration were necessary throughout the course of data collection. Speech stimuli were routed via a single TDH-50P headphone from the audiometer to the subject.

Each subject was given three test sessions on different days. For each session, the subject listened to word recordings of 36 words in a randomly-determined sequence, hence, all 108 words will be measured within the three

Table 1 Descriptive statistics for age (in years) and pure-tone thresholds (dB HL) for the 20 Singapore Mandarin speakers used for list development.

Characteristic	Mean	Minimum	Maximum	Standard Deviation
Age(years)	19.9	19	25	1.8
Pure Tone Threshold (dB HL)				
250 Hz	12.0	5.0	15.0	3.4
500 Hz	11.8	5.0	15.0	3.4
1000 Hz	7.3	0.0	15.0	4.4
2000 Hz	4.0	-5.0	15.0	6.8
4000 Hz	3.3	-5.0	15.0	6.1
8000 Hz	12.0	5.0	15.0	3.4
PTA ^a	10.3	3.3	15.0	3.1

^a PTA arithmetic average of thresholds at 500, 1000 and 2000 Hz.

sessions. Each disyllabic word was presented at 24 different intensity levels, beginning at -10 dB HL and ascending to 26 dB HL in 2 dB increments. There was a practice list in the beginning of the session in order to allow subjects to become familiar with the test. Prior to evaluation, the following instructions were given in Singaporean Mandarin:

You will hear disyllabic words presented at varying loudness. At the very soft levels it may be difficult for you to hear the words. Please repeat the words that you hear. If you are unsure of a word, you are encouraged to guess. Do you have any questions?

Subjects repeated words verbally which were scored as being correct or incorrect by a native judge who spoke Singaporean Mandarin. Two native judges were present at the test as to verify that the subject's response was correct. A word is deemed correct only if he/she was able to completely correctly identify both syllables of the disyllabic word. Each subject was allowed to have rest periods during each test session.

After the raw data were compiled, logistic regression was used to obtain regression slopes for each of the 108 disyllabic words. The 108 words were then ranked from steep to shallow slopes. The 100 words with the steepest logistic regression slopes were then arranged into ten balanced lists of 10 words each. Each of the lists was counterbalanced by random block assignment. Thus the first 10 words from the rank-ordered list of 100 words were randomly assigned, each to one of the ten lists. This process was repeated until each list contained 10 words. The eight words that were left out not only had relatively shallow slopes, three of them also had mean thresholds that were considerably different from the selected 100 words. Single factor ANOVA analysis was carried out to check for list equivalency across the ten word lists before subjecting the words to intensity level adjustments using Audacity.²⁶ To ensure greater homogeneity across the lists, the intensity of each disyllabic list was digitally adjusted so that the 50% threshold of each list was equal to the mean threshold of the participants (10.3 dB HL).

The compiled wordlists were saved onto 10 separate tracks on the CD. A 1000-Hz calibration tone was inserted at the beginning of the recording with duration of 30 s. At the start of each track, a pre-recorded message introducing the wordlist number in mandarin was inserted. A 3 s interval between words was inserted with no carrier phase before each word. The presentation level of speech materials via the headphone from the audiometer was calibrated at 20 dB SPL, which equates to 0 dB HL for speech according to ANSI, 1989.¹⁷

4 Performance-intensity (PI) function of SC-10 wordlist on normal-hearing subjects

The wordlist was tested on 25 young, normal hearing adults who are native users of Singapore Mandarin. Diagnostic audiometric tests, consisting of pure tone audiometry, tympanometry and otoscopy, were performed on the participants. These subjects all have hearing sensitivity less than or equal to 20 dB HL at 500, 1000 and 2000 Hz, and normal middle ear function. Testing was done in a sound

treated room having ambient noise levels within the prescribed limits (ANSI, 1996) and calibrated audiometer (Madsen Itera II) with TDH-49 earphones.

Speech recognition threshold (SRT) and word recognition scores (WRS) were measured on the 25 subjects based ASHA guidelines.² The order in which wordlist were used to test individual subjects is balanced to ensure equal distribution of the wordlists at different testing intensity levels. The sequence in which the word lists were presented to each subject was predetermined. Subjects were allocated a particular sequence based on their unique sequence number, i.e., the 5th subject in the study would be allocated the 5th sequence. This is also to ensure that each wordlist was used the same number of times for the speech test among subjects. As much as possible, subjects were not tested on the same list more than once, except for reusing the lists that were presented sub-threshold at SRT -5 dB and SRT -10 dB to test for SRT $+30$ dB and SRT $+40$ dB respectively.

The first word was presented at 30 dB above the subject's PTA threshold of 500, 1000 and 2000 Hz. Presentation level was reduced by 10 dB for every correctly repeated word. For incorrectly repeated word, the subsequent word was presented at the same level, and presentation level increased by 10 dB after two consecutive incorrectly repeated words. The lowest presentation level that can be correctly repeated was taken as the "base level" which was used in the second part to determine the Speech Recognition Threshold (SRT). Using the same list, 5 words were presented at this base level to record the number of correctly repeated words. Another 5 words were presented at 5 dB softer and scored according to the number of correct repetitions. This step was repeated until no more words could be repeated correctly. SRT was calculated using the Spearman-Kärber equation.²

$$\text{SRT}(\text{dB HL}) = i - r + 4.5$$

In this equation, i refers to the base level and r is the number of correct response obtained in total.

PTA (0.5, 1 and 2 kHz) and the SRT calculated was subjected to Shapiro-Wilk test to check for normal distribution, before Pearson correlation coefficient test to examine the correlation between PTA and SRT.

Word recognition score were computed at 6 levels (SRT-10, SRT-5, SRT, SRT+5, SRT+10, and SRT+20 dB) using a word list each. All 10 words within each list were presented at the same level and the percentage of correctly repeated words was recorded to obtain word recognition score to plot a performance-intensity (P-I) curve for each subject, using the following Sigmoid function,

$$f(x) = \frac{1}{1 + e^{-k(x-x_0)}}$$

In this equation, x refers to the intensity level, k refers to the steepness of the curve, and x_0 refers to the curve's mid-point intensity level.

To investigate the validity of the wordlists in speech audiometry, the presentation level at which the subject obtained a WRS of 50% and the gradient of linear portion of the curve can be obtained for comparison with mean SRT and slope of other wordlists of other languages (English and Standard mandarin) respectively.

Test-retest reliability was carried out on all subjects at three levels (SRT-5, SRT, and SRT+5 dB) using a different word list each. Shapiro–Wilk test was carried out to establish normality. Paired *T*-test was carried out to compare if there is significance difference between word recognition score when the same participant is tested again at the same intensity level using a different wordlist.

Results

Phase 1: Equivalence analysis of Singapore Mandarin disyllabic word list

With the use of equation (1), the regression slope and intercept for each of the 108 disyllabic words can be calculated.

$$\ln \frac{P}{(1-P)} = a + bi \tag{1}$$

In Equation (1), *P* is the proportion of participants who correctly identify the word at any given intensity level, *a* is the regression intercept, *b* is the regression slope, and *i* is the intensity level in dB HL. These values are used to derive a prediction on the theoretical percentage of correct recognition at any specified intensity level. Using equation (2), the percentage of correct recognition was calculated for each disyllabic words for a range of -15dB HL to 23 dB HL in 1 dB increment.

$$P(i) = P = \frac{1 - e^{(a+bi)}}{1 + e^{(a+bi)}} \times 100 \tag{2}$$

In Equation (2), *P* is the percentage of correct recognition at intensity level *i* in dBHL, *a* is the regression intercept, *b* is the regression slope.

The 108 words were ranked from steep to shallow slopes before arranging the top 100 words with the steepest

logistic regression slopes into ten 10-words word lists. The ten equivalent SC-10 disyllabic word lists are shown in Table 2 with the same ten lists in Hanyu Pinyin Romanization are presented in Table 3 and English definition in Table 4. The data for the threshold, slope at threshold, and slope from 20 to 80% for each list are presented in Table 5. The means for the slopes in the psychometric functions for the ten lists are between 8.0%/dB and 9.8%/dB.

The single factor ANOVA analysis results showed no significant difference in both the mean thresholds ($f = 0.066$, $df = 9$, $p = 0.99$) and regression slopes ($f = 0.0077$, $df = 9$, $p = 0.99$) between the ten word lists. Although there was no statistically significant difference among the ten lists, intensity level adjustments were digitally completed using Audacity in order to further improve the auditory homogeneity of the lists.²⁶ It should be noted that only minor (-0.7 to 0.2 dB) adjustments to the words in the lists were required to equate the lists. The intensity adjustments made to each wordlist are presented in Table 5.

The psychometric functions for each disyllabic word before intensity adjustment are presented in Fig. 1; the mean psychometric functions for the disyllabic lists are presented in Fig. 2. Fig. 3 contains mean psychometric functions of the disyllabic lists after intensity adjustments were made to equate performance. The predicted psychometric functions were very similar after making intensity adjustments to equate performance. From the developed speech material, a masking noise with the same word frequency was produced.

Performance-intensity (PI) function of normal-hearing subjects

The performance of the SC-10 was validated with a group of young normal hearing subjects. The details (age, gender, PTA, SRT and PTA-SRT) of the participants can be found in Table 6. From Table 6, it can be seen that there is a very

Table 2 Singaporean Mandarin disyllabic lists in rank order from steepest to shallowest slope.

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10
参加	存在	教育	特别	随着	可以	月亮	知道	朋友	可能
安全	认为	应该	时间	目标	耳朵	我们	虽然	广告	天气
偶像	问题	如果	恚气	提高	造成	能够	城市	法律	成为
告诉	开始	苹果	学习	完全	比较	呕吐	条件	工作	采取
自己	内容	重点	原价	家庭	所以	有些	另外	最后	电话
旅游	所有	今天	学生	医院	老人	地方	时候	恶人	品牌
娱乐	了解	很多	帮助	采访	非常	中国	许多	努力	出现
生活	变化	发展	下午	取得	国家	明显	能力	欢迎	欧洲
原因	产品	美食	电影	恩人	美国	或者	责任	上午	汽车
起来	哀伤	包括	人员	真正	其实	儿子	以后	文章	按照

Table 3 Singaporean Mandarin disyllabic lists in Romanised Hanyu pinyin form.

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10
cānjiā	cúnzài	jiàoyù	tèbié	suízhe	keyǐ	yuèliàng	zhīdào	péngyǒu	kěnéng
ānquán	rènwéi	yīnggāi	shíjiān	mùbiāo	èrduō	wōmen	suīrán	guānggào	tiānqì
òuxiàng	wèntí	rúguō	òuqì	tígāo	zàochéng	nénggòu	chéngshì	fǎlù	chéngwéi
gàosù	kāishǐ	píngguō	xuéxí	wánquán	bǐjiào	òutù	tiáojiàn	gōngzuò	cǎiqǔ
zìjǐ	nèiróng	zhòngdiǎn	yuánjià	jiāting	suǒyǐ	yǒuxiē	língwài	zuihòu	diànhuà
lǚyóu	suǒyǒu	jīntiān	xuéshēng	yīyuàn	lǎorén	dìfāng	shíhòu	èrén	pǐnpái
yúlè	liǎojié	hēnduō	bāngzhǔ	cǎifǎng	feicháng	zhōngguó	xǔduō	nǚlì	chūxiàn
shenghuó	biànhuà	fāzhǎn	xiàwú	qǔdé	guójiā	míngxǐǎn	nénglì	huanyíng	òuzhōu
yuányīn	chǎnpǐn	měishí	diànyǐng	ēnrén	měiguó	huòzhě	zérèn	shàngwǔ	qìchē
qīlái	āishāng	bāokuò	rényuán	zhēnzhèng	qíshí	érzi	yǐhòu	wénzhāng	ànzhào

Table 4 Singaporean Mandarin disyllabic lists in English.

List 1	List 2	List 3	List 4	List 5	List 6	List 7	List 8	List 9	List 10
Attend	Exist	Education	Special	With	Can	Moon	Know	Friends	Maybe
Safety	Believe	Should	Time	Goal	Ear	We	Although	Advertisement	Weather
Idol	Question	If	Helium	Improve	Cause	Able	City	Law	Become
Tell	Start	Apple	Learn	Complete	Comparison	Vomiting	Terms	Jobs	Take
Myself	Content	Focus	Original price	Family	And so	Some	In addition	Lastly	Phone
Holiday	All	Today	Student	Hospital	Elderly	Place	Time (moment)	Wicked person	Brands
Entertainment	Understand	Many	Help	Interview	Very	China	A lot of	Word hard	Appear
Life	Change	Progression	Afternoon	Obtain	Country	Obvious	Ability	Welcome	Europe
Reason	Product	Delicacies	Movie	Benefactor	America	Or	Responsibility	Noon	Car
Wake up	Sad	Including	Staff	Real	Actually	Son	After	Article	According to

Table 5 Mean performance of Singaporean Mandarin disyllabic lists.

List	Regression intercept	Regression slope	Slope at 50%	SD of slope at 50%	Slope from 20% to 80%	SD of slope from 20% to 80%	Threshold (dB HL)	Δ dB (dB HL)
1	3.5099	-0.3468	8.7	2.66	7.5	2.30	10.5	0.2
2	3.5238	-0.3782	9.5	2.16	8.2	1.87	9.6	-0.7
3	3.2744	-0.3224	8.1	2.00	7.0	1.73	10.5	0.2
4	3.6219	-0.3556	8.9	1.86	7.7	1.61	10.5	0.2
5	3.3296	-0.3388	8.5	2.11	7.3	1.83	10.0	-0.3
6	3.7780	-0.3624	9.1	2.62	7.8	2.27	10.3	0.0
7	3.6596	-0.3787	9.5	1.97	8.2	1.71	10.0	-0.3
8	3.7096	-0.3932	9.8	5.09	8.5	4.41	10.3	0.0
9	3.1271	-0.3201	8.0	2.08	6.9	1.81	10.0	-0.3
10	3.0797	-0.3210	8.0	1.91	6.9	1.66	9.8	-0.5
Mean	3.4614	-0.3517	8.8	2.45	7.6	2.11	10.2	-0.1
Range	0.6983	0.0731	1.8	3.23	1.6	2.80	0.9	0.9
SD	0.245608	0.026407	0.7	NA	0.6	NA	0.3	0.3

good agreement between the SRT and PTA threshold. The difference between PTA and SRT of each subject is within the clinical acceptable difference of 10 dB HL. Using Shapiro–Wilk and bivariate Pearson correlation coefficient test, it was found that the PTA and SRT obtained from the participants did not differ significantly from the normal distribution (PTA: $f = 0.596$, SRT: $f = 0.110$) and have a

strong positive linear relationship ($f = 0.00 < 0.01$, $r^2 = 0.889$) respectively.

The average word recognition scores for this group of subjects at each presentation level are summarized in Table 7. Curve estimation using the sigmoid function were carried out to plot the P–I curve (See Fig. 4). Regression analysis was used to calculate the mean slope. Linear

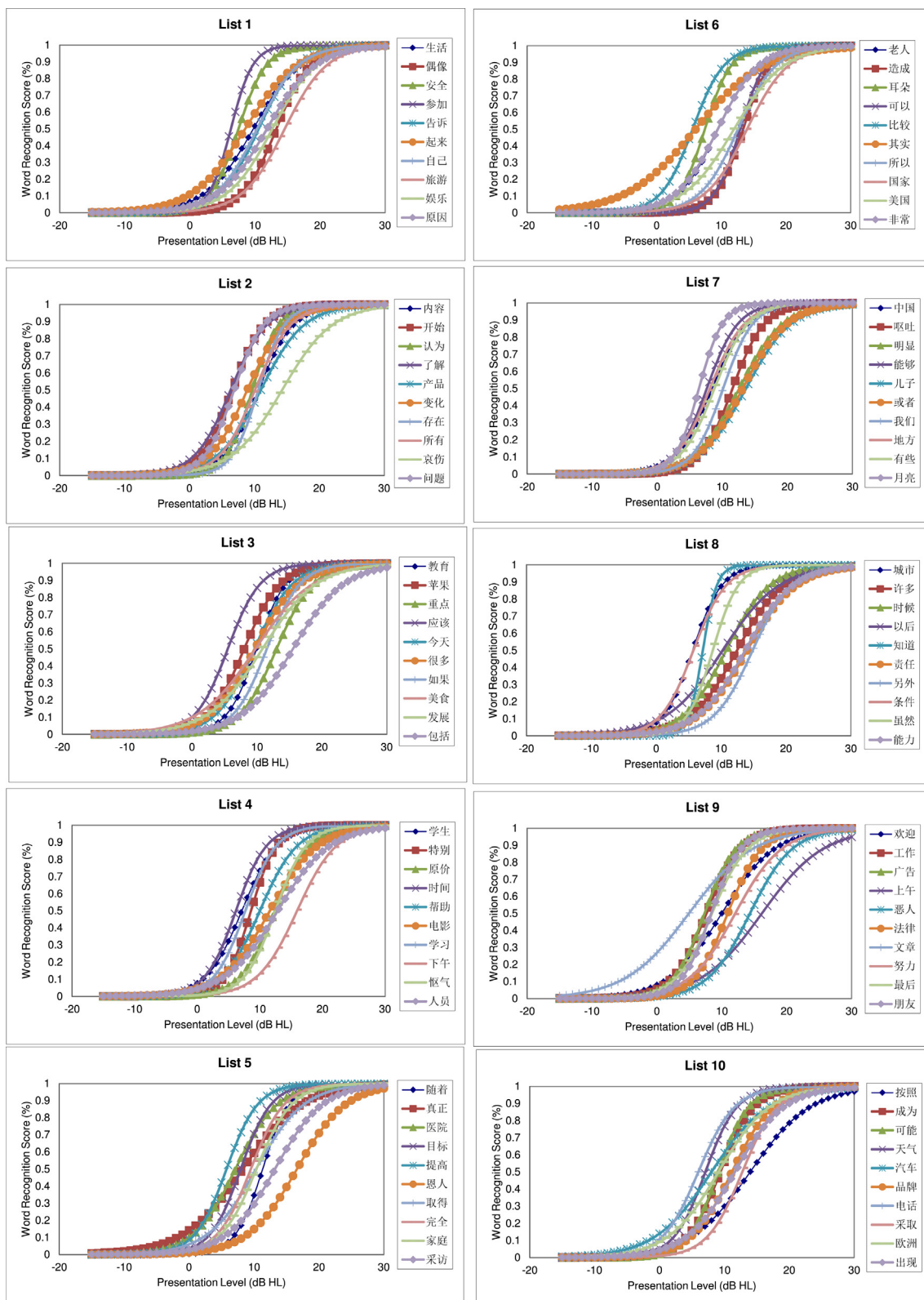


Fig. 1 Psychometric functions for the 10 Mandarin disyllabic lists.

regression analysis shows a 8.395% increase per dB in word recognition at the linear slope of the function. Quadratic regression analysis showed a high correlation coefficient ($r^2 = 0.923$) between correct score and presentation

level. The mean sound pressure level of speech for 50% word recognition score was calculated to be 9.34 dB HL. This corresponds to the mean PTA of 11 dB HL of the 25 ears.

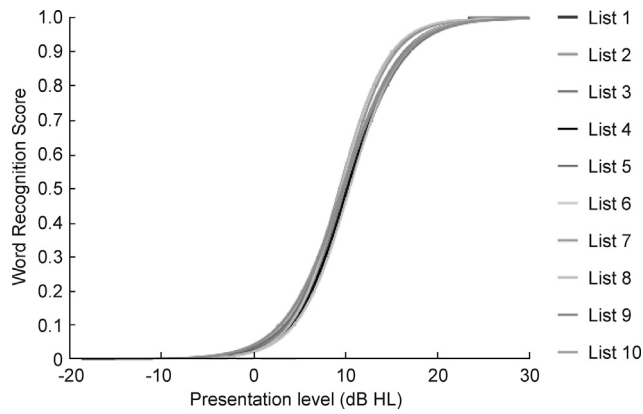


Fig. 2 Mean psychometric functions for the disyllabic lists before intensity adjustment.

Paired *T*-test results for test-retest reliability carried out on 20 out of 25 young participants at selected intensity levels (SRT-5, SRT, SRT+5) showed no significant difference between WRS when the same participant is tested at the same intensity level using a different word list (sig = 0.631, 0.982, 0.074) (See Table 8).

Discussion

1 Clinical efficiency

It was the aim of this study to develop psychometrically equivalent word lists in Singapore Mandarin that can be used for clinical speech audiometry. With the typical time allocated for speech audiometry being 30 min or less, the SC-10 wordlist created to maximise clinical efficiency and keep test time to a minimum. It was found that a complete speech audiometry test using SC-10 word list takes approximately 15 min to test one ear and 30 min to test both ears using two SC-10 word lists per ear. On the contrary, the Standard Mandarin list of 50 monosyllabic words may take at least an hour for one ear and two hours for both ears.²⁷ This is too time consuming to be used in our local clinical setting.

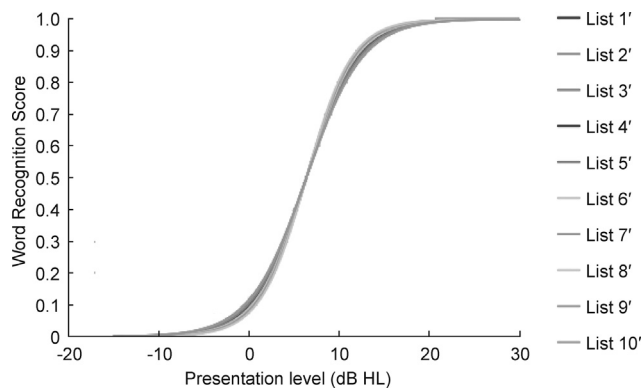


Fig. 3 Mean psychometric functions the disyllabic list after intensity adjustment.

Table 6 Descriptive statistics for age, gender, PTA, SRT and PTA-SRT for 25 normal hearing participants.

S/No	Age (years)	Gender	Ear	PTA (dB HL)	SRT (dB HL)	PTA-SRT (dB HL)
1	21	F	R	6.7	4.5	-2.2
2	23	F	R	15.0	11.5	-3.5
3	23	F	R	10.0	7.5	-2.5
4	23	F	R	15.0	12.5	-2.5
5	32	F	R	5.0	6.5	1.5
6	29	F	R	3.3	4.5	1.2
7	21	F	R	15.0	15.5	0.5
8	23	M	R	11.7	7.5	-4.2
9	23	M	L	20.0	16.5	-3.5
10	23	F	R	8.3	7.5	-0.8
11	23	F	R	13.3	9.5	-3.8
12	23	F	R	11.7	11.5	-0.2
13	23	F	R	11.7	6.5	-5.2
14	32	M	L	6.7	5.5	-1.2
15	23	M	R	11.7	10.5	-1.2
16	21	M	L	8.3	9.5	1.2
17	23	F	L	11.7	10.5	-1.2
18	21	M	R	6.7	6.5	-0.2
19	21	M	R	16.7	15.5	-1.2
20	21	M	R	8.3	5.5	-2.8
21	23	F	R	18.3	13.5	-4.8
22	21	F	R	6.7	7.5	0.8
23	22	F	L	15.0	12.5	-2.5
24	23	F	L	10.0	6.5	-3.5
25	36	M	R	8.3	8.5	0.2
Mean	23.4			11.0	9.3	-1.7
SD	4.5			4.3	3.5	2.00

2 Homogeneity of wordlist

The steeper the mean linear slope of the psychometric curve, the more homogeneous the psychometric characteristics of words.²⁸ The slopes at 50% of the psychometric curve of the 10 word lists ranged between 8.0%/dB and 9.8%/dB, with the mean slope being 8.8%/dB. Results from

Table 7 Word recognition scores at each presentation level for 25 normal-hearing subjects (25 ears).

Presentation level in relation to SRT (dB HL)	Mean	SD
-10	0.02	3.57
-5	0.29	19.19
0	0.69	15.62
5	0.88	13.47
10	0.98	3.88
15	1.00	0.00
20	1.00	0.00
25	1.00	0.00
30	1.00	0.00
35	1.00	0.00
40	1.00	0.00
45	1.00	0.00

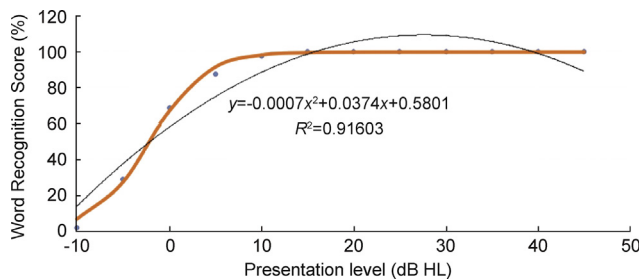


Fig. 4 P–I curve for 25 participants involved in the validation of SC-10.

the single factor ANOVA also indicated no significant differences among the 10 word lists, in terms of both mean thresholds and regression slopes. After digital adjustments, the wordlist was validated clinically on 25 normal hearing subjects and yielded a slope of 8.4%/dB.

It is noted that the mean slope at 50% for the Standard Mandarin disyllabic word list, was 6.4%/dB.²⁹ This corroborates the validity of the developed SC-10 word lists, and that the optimisation and digital adjustment were sufficient. However, when compared to the English spondees with the mean slope at 50% of the P–I function for CID W-1 at 10.0%/dB,³⁰ it indicates that further optimisation can be conducted, such as at a word level instead of a word list level. Since the SC-10 developed for Singapore Mandarin speech audiometry tests is comparable to other disyllabic Mandarin word lists developed for speech audiometry, the current optimisation is considered sufficient and this word-basis optimisation can be considered for future work.

The results on 25 normal hearing subjects showed a very good agreement between the PTA thresholds and the SRT, as well as good test-retest reliability, showing that the results obtained are accurate and reliable. The authors acknowledge that making the speech discrimination lists homogeneous for normally hearing subjects does not insure that list equivalency will be maintained for hearing-impaired subjects. Therefore, future research with varying types and degrees of hearing impairment the need to validate the word lists on a larger population of Singapore Mandarin speakers, including those with various types and configurations of hearing impairments.

Table 8 Paired *T*-test results of test-retest reliability for 20 out of 25 young participants recruited for validation of SC-10, over selected levels (SRT-5, SRT, SRT+5).

Characteristic	Mean	Paired <i>T</i> -test	
		Df	Sig. (2-tailed)
SRT-5	27.25	19	0.632
SRT-5 (retest)	28.75	19	
SRT	67.75	19	0.682
SRT (retest)	66.00	19	
SRT+5	89.25	19	0.074
SRT+5 (retest)	85.75	19	

Conclusion

To address the current lack of suitable Mandarin speech audiometry materials in Singapore clinics, the authors have developed 10 lists of 10 disyllabic words that have been found to be familiar to native Singapore Mandarin speakers. These words have been digitally recorded into individual tracks within a CD, and the lists have been designed to be psychometrically equivalent and homogeneous. The materials are the first of its kind in Singapore, and would enable Mandarin speech audiometry to be conducted in Singapore clinics. The materials also constitute an important first step towards the development of other speech tests, such as sentence tests and speech-in-noise tests, in order to ultimately achieve a comprehensive battery of Mandarin speech audiometry tests locally.

Declaration of Competing Interest

The authors have no conflict of interest to declare.

CRedit authorship contribution statement

Gary Jek Chong Lee: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft.
Steven Lock Hey Lee: SC-10 Wordlist, Validation, Writing - review & editing.

Acknowledgements

The authors would like to thank Edmund Yap Ming Sheng, Lim Si Hao, Nicholson Chang Yue Sin, Lu Yu Min, Evangeline Wong Bao Xian, Pichanida Keong (then final-year students at Temasek Polytechnic), Dr. Peng Rui, Kimberly Soh and Lee Si Ting (National University of Singapore) for their invaluable inputs and contributions to this study.

References

- Konkle DF, Rintelmann WF. Introduction to speech audiometry. *Principles of Speech Audiometry*. 1983:1–10.
- American Speech-Language-Hearing Association. *Determining Threshold Level for Speech*; 1988. <https://www.asha.org/policy/gl1988-00008.htm>.
- Hall JW. Diagnostic applications of speech audiometry. *Semin Hear*. 1983;4:179–203.
- Talbott RE, Larson VD. Research needs in speech audiometry. In: *Seminars in Hearing*. vol. 4. Thieme Medical Publishers, Inc.; 1983:299–308.
- Joint Audiology Committee on Clinical Practice Algorithms and Statements. *Audiology Clinical Practice Algorithms and Statements*. USA: Audiology Today; 2000.
- Audiology Australia. *Professional Practice Standards- Part B Clinical Standards*. 2013.
- Nissen SL, Harris RW, Jennings LJ, Eggett DL, Buck H. Psychometrically equivalent Mandarin disyllabic speech discrimination materials spoken by male and female talkers. *Int J Audiol*. 2005;44:379–390.
- Ma DY, Shen H. *Handbook of Acoustics*. Beijing, PRC: Science; 2004.

9. Ma XB, Zhao Y. *The Textbook of Mandarin*. Guangzhou, China: Jinan University Press; 2001.
10. Ma X, McPherson B, Ma L. Chinese speech audiometry material: past, present, future. *Hear Balance Commun*. 2013;11:52–63.
11. Soh W. *Validation of Mandarin Speech Audiometry Materials in Singapore*. NUS; 2017. from <http://medicine.nus.edu.sg/dgms/audiology/Research.shtml>. Accessed December 25, 2019.
12. Singapore Department of Statistics. *Table 25 Resident Population Aged 15 Years and over by Language Literate in, Age Group, Sex and Ethnic Group*. General Household Survey; 2015:2016.
13. Lee JC, Danker AN, Wong YH, Lim MY. Hearing loss amongst the elderly in a southeast asian population - a community-based study. *Ann Acad Med Singapore*. 2017;46:145–154. <http://www.ncbi.nlm.nih.gov/pubmed/28485462>.
14. Lock G. Aspects of variation and change in the Mandarin Chinese spoken in Singapore. *Aust J Linguist*. 1989;9:277–294.
15. Ng BC. A study of the variable/sh/in Singapore Mandarin. *Pacific Linguistics. Ser A Occas Pap*. 1985;67:31–37.
16. Chua CL. *The Emergence of Singapore Mandarin: A Case Study of Language Contact*. University of Wisconsin-Madison; 2003.
17. Brandy WT. *Speech Audiometry*/Katz J. *Handbook of Clinical Audiology*. 5th ed. Philadelphia, USA: Lippincott Williams & Wilkins; 2002.
18. 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. <http://www.ncbi.nlm.nih.gov/pubmed/21319936>.
19. Ramkissoon S, MacArthur I, Ibrahim M, de Graaf H, Read RC, Preston A. A qPCR assay for Bordetella pertussis cells that enumerates both live and dead bacteria. *PLoS One*. 2020;15, e0232334. <http://www.ncbi.nlm.nih.gov/pubmed/32353041>.
20. Shi Y. The establishment of modern Chinese grammar. In: *The Formation of the Resultative Construction and its Effects* Amsterdam – Philadelphia: John Benjamins; 2002.
21. Elder TE, Davis H. The articulation function of patients with conductive deafness. *Laryngoscope*. 1951;41:891–909.
22. Lehiste I, Peterson GE. Linguistic considerations in the study of speech intelligibility. *J Acoust Soc Am*. 1959;31:280–286.
23. Martin FN, Champlin CA, Perez DD. The question of phonetic balance in word recognition testing. *J Am Acad Audiol*. 2000; 11:489–493. quiz 522 <http://www.ncbi.nlm.nih.gov/pubmed/11057733>.
24. Goh HH, Lin J, Zhao CS. *The Frequency Dictionary of Daily Chinese Words Encountered by Singapore Students*. Singapore: NTU-SCCL Press; 2013.
25. Wilson RH, Strouse A. Psychometrically equivalent spondaic words spoken by a female speaker. *J Speech Lang Hear Res*. 1999;42:1336–1346. <http://www.ncbi.nlm.nih.gov/pubmed/10599616>.
26. Team AD. *Audacity (Version 1.2. 6)*; 2008. Available [audacity.sourceforge.net/download](http://sourceforge.net/download).
27. Han D, Wang S, Zhang H, et al. Development of Mandarin monosyllabic speech test materials in China. *Int J Audiol*. 2009;48:300–311. <http://www.ncbi.nlm.nih.gov/pubmed/19842805>.
28. Wilson RH, Carter AS. Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *J Am Acad Audiol*. 2001;12:7–14. <http://www.ncbi.nlm.nih.gov/pubmed/11214979>.
29. 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;46:719–731. <http://www.ncbi.nlm.nih.gov/pubmed/18049961>.
30. Young Jr LL, Dudley B, Gunter MB. Thresholds and psychometric functions of the individual spondaic words. *J Speech Hear Res*. 1982;25:586–593. <http://www.ncbi.nlm.nih.gov/pubmed/7162160>.

Edited by Jie Gao