Development and Evaluation of the Mandarin Quick Speech-in-Noise Test Materials in Mainland China

Rui Zhou[1], **Hua Zhang**[2*], **Shuo Wang**[2], **Jing Chen**[2] and **Dandan Ren**[3]

[1]*Department of Otolaryngology Head and Neck Surgery, Peking University First Hospital, Beijing, China*

[2]*Clinical Audiology Center, Beijing Tongren Hospital, Capital Medical University, Beijing Institute of Otolaryngology, Beijing, China*

[3]*Chinese AIDS Center for the Disabled, Beijing, China*

*****Corresponding author:** Hua Zhang, Clinical Audiology Center, Beijing Tongren Hospital, Capital Medical University, Beijing Institute of Otolaryngology, Beijing, China, Tel: 86-10-5826 5815; E-mail: a-zhang@263.net

Abstract

Objective: To develop and evaluate the Mandarin Quick Speech-in-Noise (M-Quick SIN) Test materials in mainland China.

Design: Four parts were included in the experiment to (1) develop sentence materials and select equivalent sentences, (2) evaluate the reliability of the lists we grouped afterwards, (3) discuss the formula of SNR loss fitted for M-Quick SIN, and (4) quantify the classification of SNR loss among normal-hearing and hearing-impaired people. 132 normal-hearing and 30 hearing-impaired subjects were participated in the experiment.

Results: A 300 sentence corpus was established and 78 sentences with better homogeneity were selected from it. After the equivalence and the test-retest reliability was established for the group materials, 11 equivalent lists for research and clinical use were chosen. The SNR-50 value for these sentences was -2 dB for normal-hearing people, and the formula was defined as "SNR loss=24.5-correct words". The classification of SNR loss was preliminarily quantified as: normal (≤ -2 dB), mild (-2 to 10 dB), moderate (10 to 20 dB) and severe (≥ 20 dB).

Conclusions: The M-Quick SIN test provided us 11 equivalent test lists (each list had 6 sentences and 30 key words) and 2 practice lists for testing normal-hearing and hearing-impaired people. The normal value of SNR-50 was -2 dB SNR, and the 6 SNRs: 20, 15, 10, 5, 0, -5 dB SNR were determined to test SNR loss for the M-Quick SIN.

Keywords: Mandarin; Speech Audiometry; Noise; Sentences; Reliability; Validity

Abbreviations: M-Quick SIN: Mandarin Quick Speech-in-Noise; SNR: Signal-to-Noise Ratio; SNR-50: SNR value required by listeners to obtain 50% correct keywords; SNR loss: Signal-to-Noise Ratio loss; PTA: Pure Tone Audiometry

Introduction

People mostly work and study in noise; therefore, communication in background noise has become a basic skill for us. But for hearing-impaired people even with hearing aids, understanding speech in background noise is one of their biggest challenges. Killion and Niquette's [1] physiologic research indicated that the loss of outer and inner hair cells causes a loss of sensitivity to quiet sounds as well as a loss of sound clarity. However, most sensorineural hearing loss patients' damage mainly due to loss of outer hair cells, so "can hear" and "can understand" became two independent concepts. The routine pure tone audiometry (PTA) could only reflect the degree of loss of sensitivity in quiet [2], which could not account for the difficulty in auditory comprehension. This test lacks the ability to predict speech in noise performance, so it must be measured directly [3]. In addition, compared to the suprathreshold monosyllabic word tests in quiet which were being used, the noise test could better simulate communication environments of daily living. Due to the limitations of PTA, the Signal-to-Noise ratio loss (SNR loss) was considered to better address these problems. SNR loss refers to the increase in SNR required by a listener to obtain 50% correct words, sentences, or words in sentences, compared to normal performance. Some published reports indicated a wide range of SNR loss in people with similar pure tone hearing losses [2,4-6]. It helped to diagnose the condition of the hearing loss more objectively and generally. Results from these tests would provide guidance for amplification strategies by judging the degree of hearing loss in noise (i.e., directional or array microphones for moderate loss, and 'FM trainer' for severe loss) [7].

The Quick SIN was developed from the original Speech in Noise test (SIN), which was compiled by Killion and Villchur [8] to evaluate speech perception in noise for hearing-impaired people under aided and unaided conditions. The test did make it easy to demonstrate that proper hearing aids improve the intelligibility of low-level speech in low-level noise, and also that they neither do not degrade the intelligibility of high-level speech in high-level noise, or improve it. But it was not considered a clinically appropriate test as it was time-consuming, had low inner-list equivalency, and was too difficult for patients [9,10] revised the SIN (i.e., RSIN) to improve the sensitivity of test and added some practice lists to reduce the learning effect. But to reduce the testing forms and decrease the time it takes to perform, the Quick SIN was developed [11]. The Quick SIN's protocol involves the presentation of six IEEE sentences [12] in multi-talker babble at 6

SNRs in which change in 5 dB steps ranging from 25 to 0 dB SNR. Each sentence has five key words concatenated in proper syntactic form with subtle semantic cues creating limited contextual cues. After the evaluation of list difficulty, nine lists were proved homogeneous with a mean of 12.2 dB SNR (SD=0.5 dB) [13]. Now the test can mainly help to diagnose SNR loss, aid audibility in noise, or assess directional-mic benefit in clinics.

Mandarin is the most commonly spoken language in the world, and consists of 23 initial consonants, 38 vowels and four tones [14]. Each phoneme and tone has a particular incidence of occurrence in the language. Since Mandarin is a tonal language, consisting of four different tones, each carrying a unique meaning [15], it is very different from English. We could not apply the results of the present studies from English-speaking subjects to Chinese people directly, but require a different approach to how hearing aids are fitted for this linguistic population. Therefore, the aim of this paper was to (1) develop sentence materials and select equivalent sentences, (2) evaluate the reliability of the lists we grouped, (3) discuss the formula of SNR loss fitted for M-Quick SIN, and (4) quantify the classification of SNR loss among normal-hearing and hearing-impaired people.

Methods

Materials

To address the characteristics of the Mandarin language and the educational level of the people simultaneously, the final selective principles for the sentences were as follows: (1) limited contextual cues; (2) for adult users at junior school reading level; (3) contained at least 5 key words per sentence and were able to meet people's short term memory capacity; (4) avoided terminology and political terms; (5) considered sentence difficulty during selection process, and avoided over homogeneity simultaneously; (6) used natural, grammatical and logical sentences; (7) simple sentences with less pattern variation, declarative sentences were preferred; (8) modern terms; (9) adequate language corpus for choosing materials. According to the criteria established above, a suitable corpus (Cao and Zhang 2009) was determined from which we chose 300 sentences based on phonetic and linguistic analyses.

Key words choices were based on both characteristics of Mandarin and the listening habits of the people of China. Taking the existence of idioms and phrases into consideration, monosyllabic, disyllabic, and polysyllabic words could appear in sentence materials [16]. The final key words' selective principles were as follows: (1) including quantity of information; (2) modal auxiliaries such as '' were permitted; (3) four-character phrases and common expression such as '' and '' were permitted; (4) negative words; (5) adverbs; (6) prosodic words; (7) numeral+quantifier: chose quantifier other than numeral; (8) form words such as '' were abandoned; (9) function words and dynamic auxiliaries such as ', '' were abandoned; (10) '' in ', ', '' etc. were abandoned; (11) considering the word's difficulty during key words' selective process, and avoided over homogeneity simultaneously as well. According to the criteria established above, we selected five key words per sentence in all 300 of the selected sentences. Four primary discourses were selected and used as the noise signals, which were chosen from the official textbooks for primary and junior school.

Recording

The 300 sentences were recorded in a standard recording studio of China National Radio Station, where the ambient noise level was lower than 25 dB (A), measured with a RION NL-11 sound-level meter. The sentences were recorded using an Electro Voice RE20 microphone connected to a Lang Xun digital audio station. Audio Cut 4 software was used in the digital audio station to collect and process the speech sounds. The recorded sentences were then transferred into a TASCAM MD-801R Mk II digital recorder through a digital tuner with 16 output channels and four input channels. The recorded sentences were converted into a CD format using a TASCAM CD-RW2000 Professional CD rewritable recorder [17].

The speaker was an experienced young female Mandarin broadcaster. Before the formal recording, the sentences were sent to her so that she could be familiar with them. During the recording, the broadcaster was seated and was asked to pronounce the sentences clearly and naturally, as well as to keep the intensity of the speech sounds at a similar level. A sound engineer and an audiologist monitored the sound level and recorded the sentences using the audio station. If a mistake was made, the sentence was recorded again.

The noise stimulus used in the test was multi-talker babble with four talkers. This noise stimulus is routinely used for the Quick SIN and is similar to the procedure used by Killion [11]. The babble noise was recorded using the test materials above with another four professional broadcasters (3 females, 1 male), and then mixed together.

Finally, 30 seconds of the calibration tone (a 1000-Hz pure tone) was inserted at the beginning of the recording [18]. The speech material in each sentence was within ± 3 dB of the standard reference 1000-Hz tone. A five-second interval between sentences was inserted. Cool Edit Prof 2.1 was used to normalize the sentences and the babble noise, to make each pair time-locked, meaning that the time relationship between each sentence and its corresponding babble segment was fixed. The sentences and the babble were transferred to the same channel, with babble always appearing 2 seconds before the sentence, and ending simultaneously with the sentence. There was a 7 second interval between sentences.

Subjects

132 normal-hearing subjects aged from 18 to 26 years old participated in this study. They were all native speakers of Mandarin, and had junior schooling or more. They had never participated in this test and did not have any prior knowledge of this experiment. Normal hearing was defined as air-conduction thresholds ≤25 dB HL [19]. Medical histories were unremarkable for otologic or hearing disorders. Only the ears with better PTA (pure tone average) threshold were used in this study. They were divided into four groups: 30 subjects in group 1 participated in Part 1 of the study, which involved the development of sentence materials and the selection of equivalent sentences. 39 subjects in group 2 participated in Part 2 of the study, which involved the evaluation of the reliability of the lists we grouped. 33 subjects in group 3 participated in Part 3 of the study, which involved the discussion of the formula of SNR loss fitted for M-Quick SIN. Another 30 subjects in group 4 participated in Part 4 of the study, which involved quantification of the classification of SNR loss among normal-hearing and hearing-impaired people

30 subjects aged from 38 to 75 years who were native speakers of Mandarin also participated in part 4. They had symmetrical, high-

frequency, sensorineural hearing losses. The selection criteria included the following: (1) a threshold at 500 Hz of ≤30 dB HL; (2) a threshold at 1000 Hz of ≤40 dB HL; (3) thresholds from 2000-8000 Hz ≥40 dB HL; (4) air-bone gaps of ≤10 dB. [20]. Only the ears with better PTA threshold were used in this part (18 left, 12 right). The average PTA threshold at 500, 1000, 2000 and 4000 Hz for these subjects ranged from 25 to 55 dB HL, with a mean of 37.1 dB HL

Procedures

The test was conducted in a sound-treated booth in Clinical Audiology Center of Beijing TongRen Hospital which met ANSI standards for ambient noise levels [21]. The materials were routed through a calibrated audiometer (GSI-61) with Cool Edit Prof 2.1 to TDH-39 earphones. The non-test ear was covered with a dummy earphone.

Part 1

All 300 sentences were divided into five groups randomly with 60 sentences per group (named group 1, group 2, etc). Each group was subjected to the 5 SNRs: +6, +3, 0, -3, -6 dB. The sentences were presented at the most comfortable level (MCL) for each subject, which were determined by a running speech recording. Then three practice lists were given to each subject to acquaint them with the testing environment and the procedure. After the practice, all 300 sentences (five groups) were heard by each subject in the 5 SNRs (+6, +3, 0, -3, -6 dB), meaning that each subject heard in total, 300 sentences×5 SNR=1500 sentences. The orders of the sentence (groups) presentations are listed in Table 1.

Subject	SNR(dB)				
	+6	+3	0	-3	-6
1	1 2 3 4 5	2 3 4 5 1	3 4 5 1 2	4 5 1 2 3	5 1 2 3 4
2	2 3 4 5 1	3 4 5 1 2	4 5 1 2 3	5 1 2 3 4	1 2 3 4 5
3	3 4 5 1 2	4 5 1 2 3	5 1 2 3 4	1 2 3 4 5	2 3 4 5 1
4	4 5 1 2 3	5 1 2 3 4	1 2 3 4 5	2 3 4 5 1	3 4 5 1 2
5	5 1 2 3 4	1 2 3 4 5	2 3 4 5 1	3 4 5 1 2	4 5 1 2 3
6	1 2 3 4 5	2 3 4 5 1	3 4 5 1 2	4 5 1 2 3	5 1 2 3 4
...	...	...	...	...	...
30	5 1 2 3 4	1 2 3 4 5	2 3 4 5 1	3 4 5 1 2	4 5 1 2 3
SNR: signal-to-noise ratio					

Table 1: The order of the sentences (groups) presentation.

Prior to the formal test, each subject was given instructions according to the Quick SIN manual (Etymotic Research) [22]:

'Imagine there is a woman talking to you and several other talkers in the background. The woman's voice is easy to hear at first, because her voice is louder. Repeat each sentence the woman says. The background talkers will gradually become louder, making it difficult to understand the woman's voice, but please guess and repeat as much of each sentence as possible.'

Because of the time-consuming procedure, each subject required three sessions to complete the test. Each session lasted approximately one and a half hours with a one week interval between every session. A short break was allowed during the procedure. The results were recorded by the same tester and "all-or-none" scoring method was used, which based on the number of correctly repeated key words. One point was given for each word(s) correctly repeated. If none were repeated correctly, the resulting score would be 0. Results were analyzed statistically using Statistical Package for the Social Sciences software, version 17.0 (SPSS 17.0). "None-linear curve fitting" was used to plot the P-I function (the recognition rate - SNR curve) of every sentence with Logic Curve [23]. An SNK-Q test (Student-Newman-Keuls) as a multiple comparison method was used in the analysis of variance (ANOVA).

Part 2

The retained 78 sentences from Part 1 with their time-locked babble were used in Part 2. An additional 12 sentences were chosen as practice sentences from the original 300 sentences, which were grouped into two lists. For this part, the 78 selected sentences were randomly ranked and 13 temporary groups were determined in order (i.e., group 1: sentence 1-6; group 2: sentences 7-12; … group 13: sentences 72-78). All of the sentences with the babble were corresponded with the 13 SNRs: 20, 18, 15, 13, 10, 8, 5, 3, 0, -2, -5, -7, -10 (dB SNR), respectively.

Subject	SNR(dB SNR)												
	20	18	15	13	10	8	5	3	0	-2	-5	-7	-10
1	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
2	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G1
3	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G1	G2
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
13	G13	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
14	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
26	G13	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
27	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
39	G13	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
G: represents Group; SNR: signal-to-noise ratio													

Table 2: The order of the sentences (groups) test for the subjects.

78 time-locked pairs were acquired and ordered into 13 groups, with 7-second intervals between sentences. Babble was presented 2 seconds before the sentence, and ended with sentence simultaneously. Prior to formal testing and following instruction, one of the two practice lists was chosen to familiarize each subject with the test. Presentation levels of the sentences were fixed at 65 dB SPL. In total, each subject listened to 78 sentences (13 groups) in recurrent SNRs ordered as above. The Latin Square Design method was used to balance the order of the sentences (Table 2). The test lasted for approximately 25 min for each subject. Each subject returned to the audiometric booth to take the test-retest with the equivalent lists after two weeks, the procedures were the same. The P-I function (the mean recognition rate - SNR curve) of each list was plotted with the "Non-linear curve fitting", then LSD (Least Significant Difference) method was used for Post Hoc Multiple Comparisons in ANOVA, then Paired-Sample T Test was used to analyze the test-retest results.

Part 3

The retained 66 sentences from Part 2 with their time-locked babble were used in Part 3. They were regrouped and corresponded with the 11 SNRs: 20, 18, 15, 10, 8, 5, 3, 0, -2, -5, -10 dB, respectively. The practice lists used here were the same as those in Part 2. In total, each subject listened to 66 sentences (11 lists) in recurrent SNRs ordered as above. The Latin Square Design method was used to balance the order of the lists (Table 3). The test lasted for approximately 20 min for each subject. The mean recognition rate of each subject was calculated under each SNR, then "Non-linear curve fitting" was used to plot the P-I function (the mean recognition rate - SNR curve) for each subject using Logic Curve.

Subjects	List numbers											
1	1	2	3	4	5	6	7	8	9	11	12	13
2	2	3	4	5	6	7	8	9	11	12	13	1
3	3	4	5	6	7	8	9	11	12	13	1	2
4	4	5	6	7	8	9	11	12	13	1	2	3
5	5	6	7	8	9	11	12	13	1	2	3	4
…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…
31	11	12	13	1	2	3	4	5	6	7	8	9
32	12	13	1	2	3	4	5	6	7	8	9	11
33	13	1	2	3	4	5	6	7	8	9	11	12

Table 3: The order of the lists for the subjects.

Part 4

The 11 equivalent lists evaluated from Part 2 were used as the test lists, and the 2 abandoned lists (list 6 and list 10) were used as the practice lists. The 6 sentences in each list corresponded with the 6 SNRs: 20, 15, 10, 5, 0, -5 dB, respectively. In total, each subject listened to 66 sentences (11 lists) in recurrent SNRs ordered as above. The levels of the sentences for the hearing-impaired subjects were presented at a level which was loud but OK. The mean SNR loss scores for normal-hearing and hearing-impaired subjects were determined. Then 1-sample K-S Test was used for normality test, and the overall mean for both groups of subjects were calculated.

Results

Part 1

Statistical analysis indicated that these sentences had great variability. Neither the SNR-50 values nor the slopes of the P-I functions were normally distributed (P<0.05). Some sentences were recognized correctly 100% of the time even in the most adverse SNR. Conversely, a few sentences were almost never understood correctly even in the most favorable SNR. The SNR-50 values varied from -25 to +3.75 dB SNR, and the slopes were in skewed in distribution. Given these results, the sentences with regression coefficients below 0.7 [24] and slopes that over steep were abandoned. The retained 78 sentences had good equivalence. The SNR-50 value for these 78 sentences was -2.00 ± 1.75 dB, with [-2.40, -1.60] dB at 0.95 level of confidence. Both the SNR-50 value and the slope were in normal distribution (P>0.05). Then, we brought the SNR-50 values of the 78 retained sentences to an expected value of -2 dB with Cool Edit Prof 2.1. For example, sentence 1 had an SNR of -3 dB, so the level of the babble associated with this sentence was reduced by 1 dB to produce the expected SNR-50 of -2 dB. All the readjusted sentences were also time-locked with the babble associated.

After readjusting the 78 sentences, they were found to have better homogeneity (Figure 1), with better concordant P-I functions, and were therefore used in Part 2 for further research.
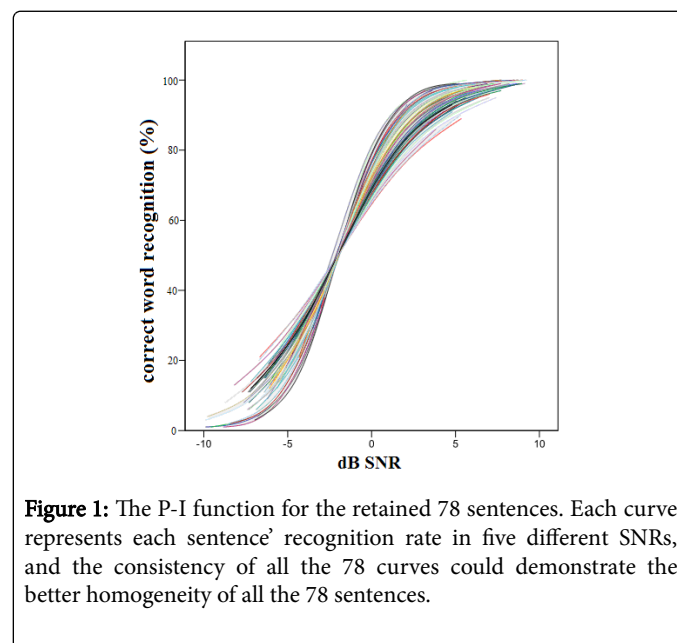
The data are illustrated in Figure 1.

Figure 1: The P-I function for the retained 78 sentences. Each curve represents each sentence' recognition rate in five different SNRs, and the consistency of all the 78 curves could demonstrate the better homogeneity of all the 78 sentences.

Part 2

Based on the 13 SNRs in this part, the mean recognition rate - SNR curves are depicted by Figure 2. Statistical analysis indicated that (1) the regression coefficients of all 13 lists were greater than 0.970, (2) and the SNR-50 values for these lists were (-2.30 ± 0.22) dB and were normally distributed (P>0.05), with [-2.35, -2.25] dB at 0.95 level of

confidence, (3) and the slopes of linear parts were (5.85 ± 0.47) %/dB and were normally distributed (P>0.05). All the results indicated that the 13 lists had better equivalence, and could be used in following research.
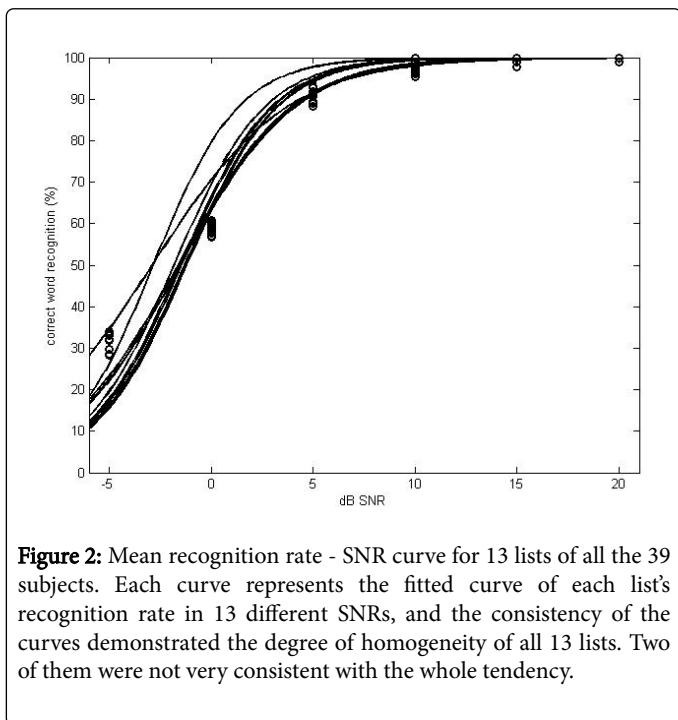
The data are illustrated in Figure 2

Figure 2: Mean recognition rate - SNR curve for 13 lists of all the 39 subjects. Each curve represents the fitted curve of each list's recognition rate in 13 different SNRs, and the consistency of the curves demonstrated the degree of homogeneity of all 13 lists. Two of them were not very consistent with the whole tendency.

Based on the 13 equivalent lists, the difference values of each pair of test-retest lists were calculated. All the data showed normal distribution but ANOVA showed discrepancy (P<0.05) (Table 4). Then we used LSD method for Post Hoc Multiple Comparisons, and found list 10 was in heterogeneity of variance with other 12 lists (P<0.05).

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	91.507	12	7.617	2.997	0.002
Within Groups	165.192	65	2.541		
Total	256.598	77			
ANOVA: analysis of variance.					

Table 4: ANOVA results.

Paired-Sample T Test showed no significant differences between re-test and initial test values, except list 6 (P<0.05, Table 5). Synthesized all the analysis above, we initially chose 11 lists (list 1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13) with better reliability for following research, and all of the 11 lists were used in Part 3.

List number	A-B	P	List Number	A-B	P
1	0.13 ± 1.16	0.802	7	-0.99 ± 0.94	0.050
2	-0.01 ± 0.41	0.969	8	0.07 ± 0.97	0.860
3	-0.11 ± 0.82	0.749	9	0.32 ± 0.93	0.436
4	-0.13 ± 0.88	0.736	11	0.20 ± 0.98	0.644
5	0.53 ± 0.93	0.222	12	0.09 ± 1.34	0.871
6	0.85 ± 0.68	0.028	13	-0.09 ± 0.95	0.832
Note: A: re-test; B: initial test; M ± SD: mean ± standard deviation					

Table 5: Comparison of recognition rate between re-test and initial test for the subjects (%, M ± SD).

Part 3

After analysis of the data from all 33 subjects, the mean recognition rate - SNR curve was plotted (Figure 3), from which the following observations could be made: (1) the SNR-50 value for these sentences was -2.24 dB and was in accordance with the result in Part 2 [-2.40, -1.60 dB]. (2) 100% recognition rate appeared at less than 10 dB SNR for normal-hearing subjects.
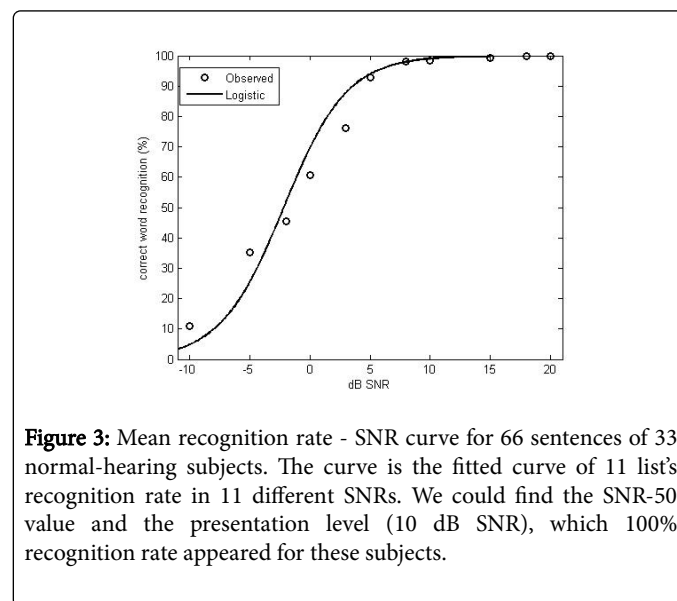
The data are illustrated in Figure 3

Figure 3: Mean recognition rate - SNR curve for 66 sentences of 33 normal-hearing subjects. The curve is the fitted curve of 11 list's recognition rate in 11 different SNRs. We could find the SNR-50 value and the presentation level (10 dB SNR), which 100% recognition rate appeared for these subjects.

The reconfirmation of -2 dB SNR as SNR-50 could better illustrate the repeatability of our sentence materials. McArdle [13] had proved an 8.7 dB difference in performances between listeners with and without hearing loss. Therefore, considering the universality of the formula between normal-hearing and hearing-impaired people, 20, 15, 10, 5, 0, -5 dB SNRs were chosen as the 6 SNRs for the following research, and 20 dB was chosen as the highest presentation level in the formula, which was written as "SNR loss=24.5-correct words". All of the 66 sentences and the 6 SNRs were used in Part 4.

Part 4

Mean SNR loss scores and standard deviation of each list for normal-hearing and hearing-impaired subjects were listed in Table 6 and Table 7, respectively, and all the data showed normal distribution. The overall mean of SNR loss scores were 0.60 ± 0.92 dB and 10.55 ± 0.77 dB for the 2 groups of subjects. Further exploration indicated the scores of the normal-hearing subjects ranged from -2.5 to 3.5 dB,

whereas the range was 0.5 to 21.5 dB for the subjects with hearing loss. Synthesized the results of both groups of subjects, -2 dB and 20 dB were considered as the normal upper limit and the abnormal lower limit, and the difference between the mean scores, 10 dB, as the boundary between mild loss and moderate loss. The classification of SNR loss was considered as: normal (≤ -2 dB), mild (-2 to 10 dB), moderate (10 to 20 dB) and severe (≥ 20 dB).

Lists	M ± SD	Lists	M ± SD
1	0.67 ± 1.42	8	0.73 ± 1.28
2	0.60 ± 1.32	9	0.50 ± 1.14
3	0.63 ± 1.20	11	0.60 ± 1.47
4	0.73 ± 1.36	12	0.53 ± 1.30
5	0.43 ± 1.28	13	0.60 ± 1.83
7	0.63 ± 1.33		
M ± SD: mean ± standard deviation.			

Table 6: SNR loss scores of each list for 30 normal-hearing subjects.

Lists	M ± SD	Lists	M ± SD
1	11.20 ± 5.72	8	10.70 ± 5.67
2	9.87 ± 6.36	9	10.30 ± 5.62
3	11.80 ± 5.31	11	9.70 ± 5.81
4	10.43 ± 5.84	12	10.43 ± 5.97
5	10.90 ± 5.67	13	11.50 ± 5.18
7	9.27 ± 6.31		
M ± SD: mean ± standard deviation.			

Table 7: SNR loss scores of each list for 30 hearing-impaired subjects.

Discussion

Understanding speech in background noise is the primary goal for hearing aids users, emphasizing the need for outcome measurements that assess speech-in-noise capabilities. Because outcome measurements could evaluate the effectiveness of intervention, they could be used to identify individuals who have difficulty understanding speech in noise, and ultimately describe the amount of difficulty and subsequent benefit provided by amplification [25]. A questionnaire study on evaluating the effectiveness of hearing aids indicated that many subjective feelings such as personal image, service and cost, and complexity of operation, determined the consumers' attitude for continuing use to some extent [26]. These evaluations, however, could not provide timely personal data and could increase the psychological burden for those who were not smooth hearing aid users. Therefore, objective tests before the fitting of hearing aids are crucial. Besides, speech in noise test materials are more similar to our verbal communication, containing natural and dynamic characteristics, which enables patients to be tested using multiple target words in a short amount of time [27]. These results can better reflect and evaluate people's communication abilities in real-world situations (i.e. noisy environments). As a tonal language, Mandarin is very different from English and other languages, so in the first parts of this study, 78 M-Quick SIN sentence materials were developed for Mandarin-speaking people.

The evaluation of reliability is an important part of the standardization of speech audiometry, which concerns the extent to which measurements are repeatable by the same individual using the same measures of a particular attribute, by the same individual using different measures of the attribute, or by different people using the same measures of the attribute without the interference of error [28]. Reliability consists of list equivalence and test-retest reliability. It represents the consistency of results among multiple lists, and the stability of results between initial and repeated tests, respectively. The reported test-retest evaluations were mostly based on the better equivalent word or sentence lists [9], so we conducted two experiments in Part 2 orderly and respectively. An effective method for equivalence evaluation was based on the consistency of SNR-50 and the slopes at those points. Figure 2 shows a cluster of functions with accordant tendency, which reflects the better equivalence of 13 lists. The SNR-50 values better agreed with those in Part 1 (-2.00 dB SNR) and Part 3 (-2.24 dB SNR). It was reported that the instantaneous slope at the 50% correct point provided an approximation of the linear slope of the function over the 20% to 70% to 80% correct points [29]. For the limited independent variable (SNR) in our experiment, we conducted the linear analysis from -5 to 5 dB SNR uniformly, which mostly included 25% to 95% correct word recognition. Both the SNR-50 values for these lists (-2.30 ± 0.22 dB) and the slopes of linear parts (5.85 ± 0.47 %/ dB) were normally distributed (P>0.05). These two parameters accounted for better equivalency for their respective consistency [29].

We used equivalent speech materials to evaluate the effectiveness of auditory rehabilitation by comparing the difference in speech audiometry at different times. Therefore, various test errors should be avoided. When evaluating test-retest reliability, the influence of outside variables should be monitored. The same conditions should be used for each subject [30]: same locations, same lists, same SNRs and so on. In addition, the test administrator should treat the subjects in the same manner among each test, so the instructions prior to formal test, and the subjects' physical and mental state should be consistent, as well. Table 4 indicated that a difference exists among 13 lists (P<0.05), and combined with LSD method for Post Hoc Multiple Comparisons, list 10 was eliminated (P<0.05). Table 5 showed the comparison among the difference values between retest and initial tests, and list 6 showed discrepancies with other 11 equivalent lists. Once the validity and reliability of the test had been established, the users could feel more confident regarding the sensitivity of the instrument (Marshall 1997).

SNR loss could be regarded as the difference between the test subjects' threshold and average-normal threshold. More precisely, SNR loss was equal to the test subjects' SNR-50 in dB minus the average-normal SNR-50 in dB. The SNR-50 was determined with a formula that included the highest presentation SNR (i.e., the lowest SNR for total recognition), the attenuation step size, and the number of correct responses. The Quick SIN manual referred to the computation as the Tillman-Olsen method [31] that was shown by Wilson et al. [32] to be a long-standing statistical precedent, the Spearman-Kärber equation [33], and chose 25, 20, 15, 10, 5 and 0 dB SNR to provide SNR-50 scores. In addition, Killion et al. have determined that the average recognition performance of a group of listeners with normal hearing on the Quick SIN to be 2 dB SNR, so once the number of correct words on a Quick SIN list was entered in the equation, the SNR loss was

easily computed by subtracting the total number of correct words from 25.5 dB SNR (i.e., SNR loss=25 dB SNR+5 dB/2-2 dB - correct words=25.5 - correct words). Since Mandarin is very different from English, we should determine the formula fitted for the M-Quick SIN. We preliminarily proved -2 dB as the average-normal SNR-50 in Part 1, and should find out the highest presentation SNR accordingly. Figure 3 showed -2.24 dB as SNR-50 after the curve-fitting of 33 subjects under 11 SNRs, the reconfirmation of -2 dB SNR as SNR-50 could better illustrate the repeatability of our sentence materials. McArdle [13] had proved an 8.7 dB difference in performances between listeners with and without hearing loss, and Figure 3 also showed that the recognition could reach 100% in less than 10 dB SNR. Therefore, considering the universality of the formula between normal-hearing and hearing-impaired people, 20 dB was considered the highest presentation SNR (the lowest SNR for total recognition) for both normal-hearing and hearing-impaired subjects, and 20, 15, 10, 5, 0 and -5 dB SNR were chosen as the 6 SNRs for the following research, and 'SNR loss=24.5 - correct words' was used as formula for M-Quick SIN.

While there was no available classification of SNR loss, and to enable the results to be used more conveniently, according to pathology research data [34], Killion and Niquette [1] suggested that a loss of 20 dB in ability to understand speech in noise excluded the patient from social conversation at parties (profound loss), and initially suggested categories for SNR loss as: mild (0-4 dB), moderate (5-10 dB), severe (11-19 dB) and profound (20 dB). Then combined with Quick SIN test, the refining classification for SNR loss (normal ≤ 2 dB, mild 3-7 dB, moderate 7-15 dB and severe ≥ 15 dB) became accepted (Etymotic Research 2001). In our experiment, the SNR loss scores of the normal-hearing subjects ranged from -2.5 to 3.5 dB, and 0.5 to 21.5 dB for the hearing impaired. So we considered -2 dB as the normal upper limit, and 20 dB as the abnormal lower limit. Table 6 and Table 7 showed the mean SNR loss scores for both normal-hearing and hearing-impaired subjects (0.60 ± 0.92 dB vs. 10.55 ± 0.77 dB), with a disparity of about 10 dB. Then 10 dB is considered the boundary between mild loss and moderate loss. A rough classification of SNR loss for the M-Quick SIN test was found to be the following: normal ≤ -2 dB means there is no SNR loss or essentially normal, and speech perception in noise is equal to or better than normal-hearing individuals; mild -2 to 10 dB means there is little SNR loss, and speech perception in noise has basically no problem; moderate 10 to 20 dB means there was significant SNR loss, and perception in noise is increasingly difficult; and severe ≥ 21 dB means more SNR loss than normal, and almost lose the capability of the perception in noise. However, the range of hearing-impaired subjects was very widely distributed (0.5 to 21.5 dB), so that a larger sample is needed in order to subdivide the level from mild to moderate.

Conclusions

The M-Quick SIN test provided 11 equivalent test lists and 2 practice lists (6 sentences and 30 key words per list) for a speech perception in background noise test. The test was time-saving, with one list taking approximately one minute to administer. The normal value of SNR-50 was -2 dB SNR, and the 6 SNRs: 20, 15, 10, 5, 0 and -5 dB SNR were determined to test SNR loss in M-Quick SIN. This study suggests that the classification of SNR loss for Mandarin speaking subjects should be as follows: normal (≤ -2 dB), mild (-2 to 10 dB), moderate (10 to 20 dB) and severe (≥ 20 dB).

References

1. Killion MC, Niquettfe PA (2000) What can the pure-tone audiogram tell us about a patient's SNR loss? Hear J 53: 46-53.

2. Killion MC (1997) SNR loss: I can hear what people say, but I can't understand them. Hear Rev 4: 8-14.

3. Killion MC (2002) New thinking on hearing in noise: a generalized articulation index. Seminars in Hearing 23: 57-76.

4. Lyregaard PE (1982) Frequency selectivity and speech intelligibility in noise. Scand Audiol Suppl 15: 113-122.

5. Dirks DD (1982) Speech discrimination ability in the hearing-impaired. In: Studebaker G, Bess F (edn.), The Vanderbilt hearing aid report-monographs in contemporary audiology, Upper Darby, PA, pp. 44-50.

6. Taylor B (2003) Speech-in-noise tests: How and why to include them in your basic test battery. Hear J 56: 40-46.

7. Killion MC (1997) The SIN report: Circuits haven't solve the hearing-in-noise problem. Hear J 51: 32-47.

8. Killion MC, Villchur E (1993) Kessler was right–partly: but SIN test shows some aids improve hearing in noise. Hear J 46: 31-35.

9. Bentler RA (2000) List Equivalency and Test-Retest Reliability of the Speech in Noise Test. Am J Audiol 9: 84-100.

10. Cox RM, Gray GA, Alexander GC (2001) Evaluation of a revised speech in noise (RSIN) test. J Am Acad Audiol 12: 423-432.

11. Killion MC, Niquette PA, Gudmundsen GI (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 116: 2395-2405.

12. Institute of Electrical and Electronic Engineers (1969) IEEE recommended practice for speech quality measurements, Appendix C. Global Engineering Documents, Boulder, CO.

13. McArdle RA, Wilson RH (2006) Homogeneity of the 18 Quick SIN Lists. J Am Acad Audiol 17: 157-167.

14. Ma DY, Shen H (2004) Handbook of Acoustics. Science Publishing Company, Beijing.

15. Ma XB, Zhao Y (2001) The Textbook of Mandarin. Jinan University Press, Guangzhou.

16. Huang BR, Liao XD (2002) Modern Chinese. Higher Education Press, Beijing.

17. Wang S, Mannell R, Newall P (2007) Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. Int J Audiol 46: 719-731.

18. Sherwood T, Fuller H (1997) Equipment for speech audiometry and its calibration. In: Martin M (edn.), Speech Audiometry (2nd ed.). London: Whurr, pp. 89-106.

19. ANSI (1996) Specification for Audiometers (ANSI S3.6-1996). New York, USA.

20. (1998) Tonal and speech materials for auditory perceptual assessment disc 2.0. Department of veterans affairs. Mountain home, VA Medical Center, TN.

21. ANSI (1999) Maximum Ambient Noise Levels for Audiometric Test Rooms. (ANSI S3.1-1999). New York, USA.

22. Etymotic Research (2001) Quick sin speech in noise test- Version 1.3. Elk Grove Village, IL.

23. Nissen SL, Harris RW, Jennings LJ (2005) Psychometrically equivalent Mandarin disyllabic speech discrimination materials spoken by male and female talkers. Int J Audiol 44: 379-390.

24. Munro BH, Visintainer MA, Page EB (1993) Statistical Methods for Health Care Research. JB Lippincott Co, Philadelphia: Pa.

25. Bray V, Nilsson M (2002) Assessing hearing aid fittings: An outcome measures battery approach. In: Valente M (2nd edn.), Strategies for selecting and verifying hearing aid fittings. New York: Thieme Publication, pp. 151-175.

26. Liu HH, Zhang F, Zhang H (2011) Evaluation of hearing aids operational complexity. Chin Sci J Hear Speech Rehab 47: 69-73.

27. McArdle RA, Wilson RH, Burks CA (2005) Speech Recognition in Multitalker Babble Using Digits, Words, and Sentences. J Am Acad Audiol 16: 726-739.

28. Bilger RC, Nuetzel JM, Rabinowitz M (1984) Standardization of a test of speech perception in noise. J Speech Hear Res 27: 32-48.

29. Wilson RH, Carter AS (2001) Relation between Slopes of Word Recognition Psychometric Functions and Homogeneity of the Stimulus Materials. J Am Acad Audiol 12: 7-14.

30. Bamford J, Wilson I (1979) Methodology considerations and practical aspects of the BKB sentence lists. In: Bench J, Bamford J, (edn.) Speech-hearing tests and the spoken language of hearing-impaired children, Academic Press, UK.

31. Tillman TW, Olsen WO (1973) Speech audiometry. In: Jerger J, (2nd edn.) Modern Developments in Audiology. Academic Press, New York, pp. 37-74.

32. Wilson RH, Morgan DE, Dirk DO (1973) A proposed SRT procedure and its statistical precedent. J Speech Hear Disord 38: 184-191.

33. Finney DJ (1952) Statistical method in biological assay. Griffen C publishing, London.

34. Schuknecht HF (1993) Pathology of the Ear (2nd edn.). Lea Febiger, Baltimore.

35. Lisa LM (1997) Audiologic evaluation and management and speech perception assessment. Singular Publishing Group Inc. San Diego, London.

36. Killion MC (1997) Hearing aids: Past, present, future: Moving toward normal conversations in noise. Brit J Audiol 31: 141-148.