

The Chinese Zodiac-in-Noise Test: An Internet-Based Speech-in-Noise Test for Large-Scale Hearing Screening

Huali Zhou,¹ Qinglin Meng,² Xiaohong Liu,³ Peina Wu,³ Shidong Shang,⁴ Wei Xiao,⁴ Yuyong Kang,⁴ Jiawen Li,¹ Yamin Wang,⁵ and Nengheng Zheng¹

Objectives: Motivated by the growing need for hearing screening in China, the present study has two objectives. First, to develop and validate a new test, called the Chinese Zodiac-in-noise (ZIN) test, for large-scale hearing screening in China. Second, to conduct a large-scale remote hearing screening in China, using the ZIN test developed.

Design: The ZIN test was developed following a similar procedure as the digits-in-noise test but emphasizes the importance of consonant recognition by employing the 12 zodiac animals in traditional Chinese culture as speech materials. It measures the speech reception threshold (SRT) using triplets of Chinese zodiac animals in speech-shaped noise with an adaptive procedure.

Results: Normative data of the test were obtained in a group of 140 normal-hearing listeners, and the performance of the test was validated by comparisons with pure-tone audiometry in 116 listeners with various hearing abilities. The ZIN test has a reference SRT of -11.0 ± 1.6 dB in normal-hearing listeners with a test-retest variability of 1.7 dB and can be completed in 3 minutes. The ZIN SRT is highly correlated with the better-ear pure-tone threshold ($r = 0.82$). With a cutoff value of -7.7 dB, the ZIN test has a sensitivity of 0.85 and a specificity of 0.94 for detecting a hearing loss of 25 dB HL or more at the better ear.

A large-scale remote hearing screening involving 30,552 participants was performed using the ZIN test. The large-scale study found a hearing loss proportion of 21.0% across the study sample, with a high proportion of 57.1% in the elderly study sample aged over 60 years. Age and gender were also observed to have associations with hearing loss, with older individuals and males being more likely to have hearing loss.

Conclusions: The Chinese ZIN test is a valid and efficient solution for large-scale hearing screening in China. Its remote applications may improve access to hearing screening and enhance public awareness of hearing health.

Key words: Hearing loss, Hearing screening, Speech in noise.

(Ear & Hearing 2024;45:451–464)

INTRODUCTION

According to the World Health Organization (2021) Hearing Report, hearing loss affects more than 1.5 billion people worldwide, or 20% of the world's population, and has become a serious global health concern (World Health Organization 2021). Hearing screening is an essential initial step in managing hearing loss because untreated hearing loss can have a major negative impact on the daily lives of

individuals affected (Arlinger 2003). According to a survey conducted in China using samples from four provinces, there are about 200 million people in China who have hearing loss of varied severity (Hu et al. 2016). With an aging population, it is projected that the prevalence of hearing loss in China will rise quickly. Despite its significance, most of the Chinese population, particularly in rural areas, lacks access to hearing screening. This study was motivated by the growing need for hearing screening in China.

Many hearing screenings for adults are survey-based and rely on questionnaires (e.g., Heine et al. 2019). However, questionnaires have been proven to have limited accuracy in detecting hearing loss (Nondahl et al. 1998; Smits et al. 2006). Pure-tone audiometry is more accurate and used as the golden standard in clinics but requires a professional audiologist, a quiet sound-proof room, and an expensive audiometer, making it prohibitively expensive for large-scale hearing screening.

Numerous efforts have been made to provide alternative hearing tests suitable for large-scale screening. One such approach involves overcoming the limitations of pure-tone audiometry by using an automated procedure with consumer electronics such as smartphones or tablets with earbuds (Guo et al. 2021; Wasmann et al. 2022). Another strategy involves using speech-in-noise tests as a screening tool. This method is based on the notion that natural speech in noise better reflects the difficulties that individuals with hearing loss face in real-world settings, as opposed to the artificial stimuli used in pure-tone audiometry (Kramer et al. 1998; Pienkowski 2017). Furthermore, the signal-to-noise ratio (SNR) resulting from speech-in-noise tests is a suprathreshold and relative measure that does not require accurate calibration of the absolute levels of speech or noise. By using speech-in-noise tests, large-scale hearing screening can be made possible using arbitrary devices at a user-adjusted volume that is adequate and comfortable.

Remote hearing screenings based on speech-in-noise tests are increasingly gaining popularity worldwide. The speech-in-noise test was first used for large-scale hearing screening in the early 2000s. In 2004, a Dutch speech-in-noise test, the digits-in-noise (DIN) test, was implemented on an Interactive Voice Response System at a telephone company (Smits et al. 2004). It was introduced as the National Hearing test in the Netherlands, and within the first 4 months, 65,924 individuals had taken the test (Smits and Houtgast 2005). Since then, multiple language versions of the DIN test have been developed and employed as screening tools in various regions around the world (refer to the review in Zokoll et al. 2013; Van den Borre et al. 2021). To date, speech-in-noise tests that incorporate various stimuli not limited to digits have been utilized for remote hearing screening across multiple platforms, including telephone, website, and

¹Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China; ²Acoustics Laboratory, School of Physics and Optoelectronics, South China University of Technology, Guangzhou, China; ³Department of Otolaryngology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China; ⁴Tencent Ethereal Audio Lab, Shenzhen, China; and ⁵School of Architecture, South China University of Technology, Guangzhou, China

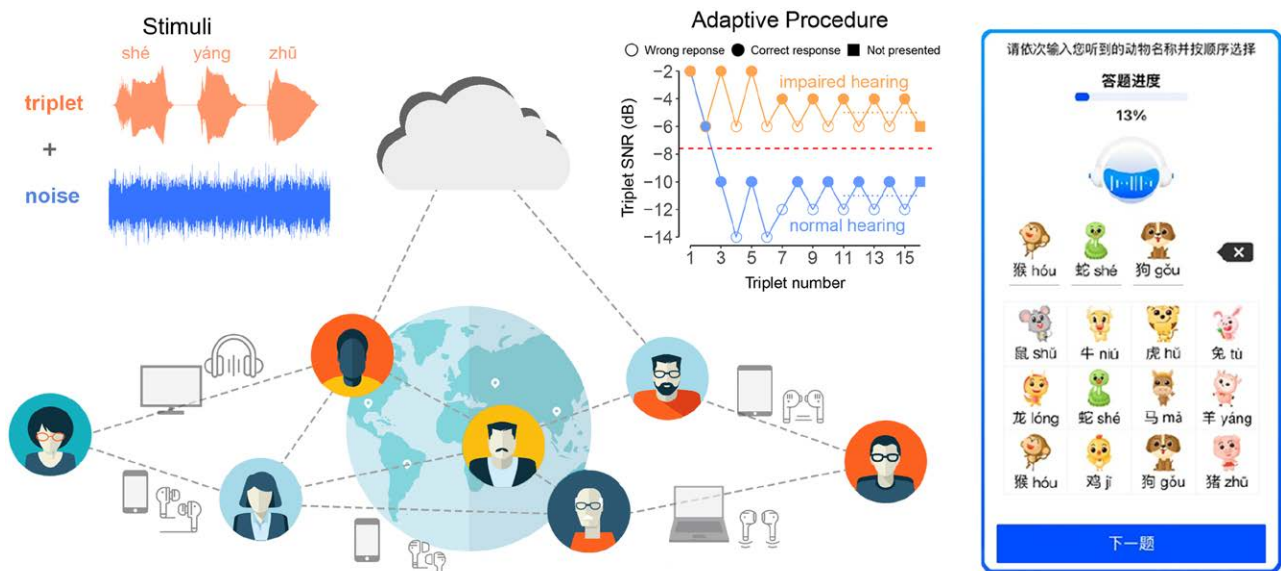


Fig. 1. Outline of the design of the ZIN test, including illustrations of the triplets-in-noise stimuli, the adaptive procedure with two demo samples for impaired and normal hearing, the application in a remote hearing screening with dissemination through social networks, and the graphical user interface (GUI). ZIN indicates Zodiac-in-noise.

mobile applications (see the review in Swanepoel et al. 2019; Almufarrij et al. 2022).

Limited data is available for large-scale hearing screening in adults in China, despite the abundance of data for newborns. Bu et al. (2011) and Hu et al. (2016) conducted the two largest studies on this topic, with sample sizes of 29,246 and 45,052, respectively, from four provinces each. However, these studies used pure-tone audiometry, resulting in high financial and time costs. To mitigate these costs, remote hearing screening based on speech-in-noise tests could be a viable option. Nonetheless, data on large-scale remote hearing screening using speech stimuli in China are currently not available. Although a Chinese DIN test and a Chinese version of the hearWHO app, both using digits in noise as stimuli, have been developed for remote hearing screening, their outcomes have been limited to a small sample size tested in a laboratory setting. To date, no data on remote screening are available (Qi et al. 2018; Fu et al. 2021). Furthermore, the Chinese digits from 0 to 9, except for 1 and 7, have distinct vowels that are easily identifiable, even if the consonants are misperceived. Consonants, in general, exhibit relatively shorter duration and higher frequencies compared to vowels, rendering them particularly susceptible to high-frequency hearing loss, which is one of the most common types of hearing loss. The distinct vowels in the Chinese DIN test weaken the ability of the test to evaluate consonant recognition (The pinyin for digits 0 to 9 are: *líng, yī, èr, sān, sì, wǔ, liù, qī, bā, and jiǔ*, respectively) (Winn and Wright 2022). This limitation could be addressed by using stimuli with more identical vowels.

The present study has two objectives. First, to develop and validate a new test, called the Chinese Zodiac-in-noise (ZIN) test, for large-scale hearing screening in China. To address the limitation of the Chinese DIN test mentioned above, the ZIN test emphasizes the importance of consonant recognition by employing the 12 zodiac animals in traditional Chinese culture. The complete development and validation of the ZIN test

are presented. Second, to conduct a large-scale remote hearing screening in China, using the ZIN test developed. With dissemination through social networks, a sample size of 30,552 participants was reached within just 1 week. The outcomes are analyzed and presented. The development and application of the ZIN test in this study aim to promote public awareness of hearing health and contribute to the prevention and management of hearing loss in China.

MATERIALS AND METHODS

Development

The ZIN test is intended for large-scale hearing screenings and has been designed to meet several specific requirements. First, it should be able to detect hearing loss using common consumer electronic devices. Second, it should be easy for the largest possible population to self-administer. Third, the test should require a low time cost. Additionally, the ZIN test is designed to emphasize the importance of consonant recognition by using speech material that has some identical vowels to prevent recognition based on solely vowels. To fulfill these requirements, the ZIN test was designed as a triplets-in-noise test, as outlined in Figure 1 and described below.

Type of Test: A Triplets-in-Noise Test • To detect hearing loss using common consumer electronic devices, the test was conceived as a speech-in-noise test, specifically a triplets-in-noise test. As mentioned in the section Introduction, triplets-in-noise tests, such as the DIN test, have been used successfully in hearing screenings in many countries. These tests have shown good performance in detecting hearing loss and do not require accurate device calibration, making them suitable for use with common consumer electronic devices. Moreover, the low cognitive and memory load of the triplets-in-noise test makes it easy for a large population to self-administer. The ZIN test can be implemented as an HTML5 webpage that supports multimedia

on cross-platform devices, including mobile phones, tablets, and computers. This approach ensures that the ZIN test is accessible to the broadest possible population and suitable for remote hearing screening with dissemination through social networks.

Speech Material: The 12 Chinese Zodiac Animals • The 12 Chinese zodiac animals (given in Table 1) were chosen as the target speech material. The selection of the 12 Chinese zodiac animals is driven by two primary considerations. First, the 12 Chinese zodiac animals are well-known and highly engaging for the vast majority of the Chinese population. Based on the Chinese lunar calendar, the Chinese zodiac follows a 12-year cycle, each year designated by a different zodiac animal, from the rat to the pig. For instance, the Chinese lunar year starting from January 22, 2023 in the Gregorian calendar was the year of the rabbit. The familiarity of the Chinese population with these zodiac animals renders them well-suited as speech test materials. In addition, the 12 Chinese zodiac animals are an important aspect of Chinese traditional culture, thus potentially enhancing participants' enjoyment and overall experience during the study.

Moreover, the use of the 12 Chinese zodiac animals was motivated by the aim to highlight consonant perception, which is more susceptible to hearing loss compared to vowel recognition (Owens et al. 1968; Bilger and Wang 1976). In the Chinese DIN test, the speech material is comprised of 10 monosyllables, namely, *líng, yī, èr, sān, sì, wǔ, liù, qī, bā, and jiǔ* in pinyin. In these 10 monosyllables, there are only 5 consonants (*l, s, q, b, and j* in pinyin), and they all have distinct vowels except for *yī* and *qī* which have no consonant (in *yī*) and an alveolar affricate (*q* in pinyin), respectively. Consequently, these monosyllables could be identified solely based on their distinct vowels, even if the consonants are misperceived. In contrast, the 12 monosyllables for the 12 Chinese zodiac animals have 9 consonants (*g, t, zh, j, sh, h, m, n, and l* in pinyin), and more syllables with identical vowels, that is, the syllables *shǔ, zhū, hǔ, tù* and *hóu, gǒu*. Therefore, listeners need to accurately perceive the consonants to avoid confusion among these syllables. For example, *shǔ* and *zhū* have the same vowel, and their consonants have the same place of articulation, differing only in the manner of articulation. Similarly, *shǔ* and *hǔ* have the same vowel, and their consonants have the same manner of articulation, differing only in the place of articulation. This highlights the importance of accurate consonant recognition, in addition to vowel recognition and lexical tones, in the recognition of zodiac animals in noise.

Material Recording and Masking Noise

The development of the ZIN test largely followed the guidelines for constructing multilingual speech tests provided by the International Collegium of Rehabilitative Audiology (Akeroyd et al. 2015), with some modifications, which will be elaborated in the following sections. In this section, speech materials were recorded and carefully selected, followed by the creation of the masking noise.

Recording • We refer to the standard ISO 8253-3 (2012) for material recording. The recording of the materials was carried out in an anechoic chamber with an ambient noise level below 20 dB(A). Recording equipment included an IEC 61672 Class 1 compliant omnidirectional microphone (MicW M215), a sound card (Fireface UC RME), and a personal computer running recording software (Audition 2019). The audio was recorded at a sampling rate of 44.1 kHz and stored as 16-bit.wav files.

The materials were recorded by four trained broadcasters, two males and two females, at a normal vocal effort and speech rate. As in some DIN tests (e.g., Smits et al. 2013), single zodiac animals were recorded for subsequent combinations into triplets for the ZIN test. Therefore, the combined triplets did not have any natural coarticulation or prosody. It was found that there were no significant differences in the results measured using triplets of digits with or without coarticulation and prosody in the DIN test (Lyzenga and Smits 2011). Each speaker made five recordings for each zodiac animal.

The recordings were then subjected to a careful selection process. Initially, the authors selected a female speaker with the most natural-sounding voice based on naturalness, articulation, voice quality, intonation, and speed of production. Ten listeners with backgrounds in electroacoustics and architectural acoustics then picked the best recordings for each zodiac animal, based on naturalness, articulation, voice quality, intonation, and speed of production, from the five recordings of the chosen speaker. The best recordings of the 12 zodiac animals (presented as a spectrogram in Fig. 2) were used as speech tokens for further development.

Masking Noise • The masking noise was created as a quasi-stationary noise with the same long-term average spectrum as the target voice. This noise was produced by shaping white noise to the long-term average spectrum of the speech tokens. The root mean square (RMS) level of the noise was then adjusted

TABLE 1. Word list: the 12 Chinese zodiac animals

No.	Chinese Character	Pinyin	International Phonetic Alphabets (IPAs)	Meaning in English
1	鼠	shǔ	ʃu ²¹⁴	Rat
2	牛	niú	niou ³⁵	Ox
3	虎	hǔ	xu ²¹⁴	Tiger
4	兔	tù	tʰu ⁵¹	Rabbit
5	龙	lóng	luŋ ³⁵	Dragon
6	蛇	shé	ʃɛ ³⁵	Snake
7	马	mǎ	ma ²¹⁴	Horse
8	羊	yáng	iaŋ ³⁵	Sheep
9	猴	hóu	xou ³⁵	Monkey
10	鸡	jī	tɕi ⁵⁵	Rooster
11	狗	gǒu	kou ²¹⁴	Dog
12	猪	zhū	tʃu ⁵⁵	Pig

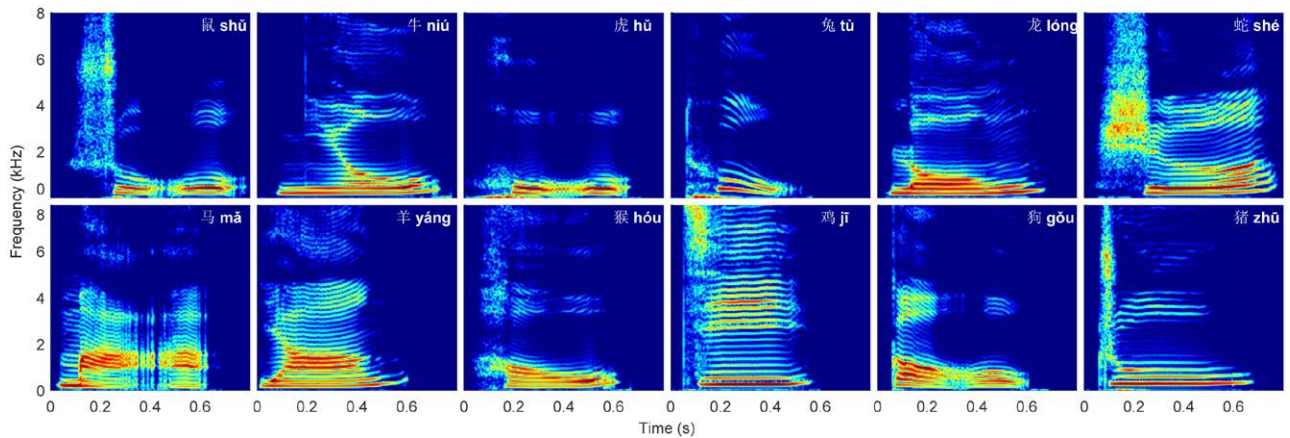


Fig. 2. Spectrograms of speech tokens (12 Chinese zodiac animals).

to match the average RMS level of the speech tokens. To prevent listeners from learning the masking noise and to ensure its freshness (Lyzenga and Smits 2011), a long noise file of 150 seconds was produced. This duration was sufficient to allow for a random selection of different noise fragments for each presentation.

Optimization

The purpose of optimization is to achieve the highest possible homogeneity of the intelligibility across the speech tokens employed. The optimization process was carried out using a two-step procedure similar to that of a previously developed smartphone-based DIN test (Potgieter et al. 2016). In the first step, the speech reception thresholds (SRTs) of the intelligibility functions of each speech token were measured in a group of normal-hearing listeners. In the second step, the level of the 12 speech tokens was adjusted to minimize the spread of SRTs.

Methods •

Participants • Twenty-four normal-hearing listeners (14 males and 10 females) with a mean age of 24.5 years (range: 19–35 years) participated in this experiment. The normal hearing status, defined as pure-tone thresholds of no more than 15 dB HL at each octave frequencies from 250 to 8000 Hz, was verified using an automated pure-tone audiometry system (Guo et al. 2021). All participants provided their informed consent before participating in all the experiments conducted in this study. This study was approved by the Ethical Review Board of Shenzhen University.

Devices • The optimization experiment was conducted in a soundproof room with a background noise level below 30 dB(A). A personal computer running custom-made software in MATLAB was used for presenting stimuli and collecting participants' responses. The stimuli were presented through a headphone set (Sennheiser HD 650) connected to the computer via an audio interface (Fireface UC RME). The system volume was set by the participant to an adequate and comfortable level.

Procedures • The experiment was conducted as a closed-set test using a one-interval 12-alternative forced choice (1I-12AFC) task. In each trial, a speech token was presented with a noise segment diotically to the participant. Fresh noise segments were randomly selected for each presentation. With a 50 ms ramp at each end, the noise began 300 ms before the speech did and continued

until 300 ms after it ended. Participants were then required to select the perceived zodiac animal from the 12 options displayed on a graphical user interface on the screen. If necessary, participants were instructed to guess. The next presentation was triggered by the response to the previous one, and the task was completely self-paced. Participants were encouraged to take short breaks whenever they wished. Before the formal test, participants went through a brief training phase in which every speech token was presented once at 0 dB and –2 dB SNR in random order. Trial-by-trial visual feedback on correctness was provided during the training phase, but not in the test phase.

To address possible differences in intelligibility among speech tokens, a pilot test was conducted with four participants to determine the specific SNR range of 10–90% intelligibility for each speech token. Subsequently, the formal test was administered to the remaining twenty participants. For the formal test, each speech token was presented in five SNRs specifically chosen based on the results of the pilot test. These SNRs are listed in Table 2. Each speech token was presented 10 times for each SNR in a random order, resulting in a total of 600 presentations for each participant (12 speech tokens × 5 SNRs × 10 presentations per token per SNR). Testing began with the easiest SNR and progressed to the most challenging. Participants set the sound output volume to a comfortable level before the test, and that level was kept constant for that participant throughout.

Psychometric Function Fitting • For each speech token, a psychometric function was fitted using raw data from all participants. The maximum likelihood method, implemented by the Wichmann-Psignifit lab's toolbox for MATLAB, was used to perform the fitting (Wichmann and Hill 2001). The following logistic function was used to fit the data:

$$SI(SNR) = 100\left(\gamma + (1-\gamma)\frac{1}{1 + e^{4s(SRT-SNR)}}\right) \quad (1)$$

where SI denotes speech intelligibility in percentages, γ represents the guess rate (set to 1/12 in this study), and SRT and s are the fitting parameters indicating the SRT and the slope of the psychometric function at SRT, respectively. Version 4 of the Psignifit toolbox was utilized in this study, and it can be downloaded from <https://www.nip.uni-tuebingen.de/research/software/psignifit.html>. The root mean squared error (RMSE) was

employed as a metric to assess the goodness-of-fit for the curve fitting for each speech token. It was calculated as the square root of the average squared difference between observed intelligibility and predicted intelligibility along the fitted curve across the range of tested SNRs.

Results • Figure 3 shows the psychometric functions that were fitted for each of the 12 speech tokens, and Table 2 presents the detailed results. Overall, the fittings are of good quality, as indicated by the average RMSE of 2.9%. To estimate the amount of level corrections required for optimization, the SRT of each speech token was subtracted from the average SRT across the 12 zodiac words, and the results were listed in the dB-diff column in Table 2. Note that dB-diff was calculated at the SRT, namely the SNR for the 50% point on the curve. It should be also noted that the 50% point on the curve corresponds to an actual intelligibility level of 54% due to a chance level of $1/12 + (1-1/12)/2 = 0.54$. Level corrections were then performed by applying a factor transformed from dB-diff to the raw audio data of the speech tokens, and the average RMS of speech tokens was matched with the noise. This enabled the psychometric curves of all speech tokens to have the same SRT. As a result, all

speech tokens had a 50% chance of being correctly recognized at the same SNR, that is, the SRT.

The speech token for the zodiac animal rooster (No. 10 in Table 2 and the leftmost curve in Fig. 3) was found to have significantly different SRT and slope values from the other tokens (the SRT was lower than the lower quartile of the set of 12 tokens by more than 1.5 interquartile range). To ensure the homogeneity of the final ZIN test, following the approach used in (Vlaming et al. 2014), it was decided to keep the rooster token while excluding its recognition results from the analysis. In other words, the response to this token would be ignored and the remaining two tokens would be scored when rooster is present in the triplet.

Implementation of the ZIN Test

Generation of Triplets • The optimized speech tokens and the constructed noise were used to generate stimuli for the ZIN test. Audio files for triplets at various SNRs were prepared and stored on the server. The SNR range was set between -20 dB and +8 dB with a step size of 2 dB. This range was chosen to be wide enough to perform adaptive SRT measurements for most individuals, regardless of their hearing ability (Smits et al. 2005; Jansen et al. 2010; Qi et al. 2018).

TABLE 2. Optimization test conditions and results

No.	Pinyin	Tested SNR (dB)	SRT (dB)	Slope (%/dB)	RMSE (%)	dB-diff
1	shǔ	-16, -14, -12, -10, -8	-11.1	13	2.3	2.7
2	niú	-17, -15, -13, -11, -9	-12.4	16	2.1	1.4
3	hǔ	-12, -10, -8, -6, -4	-9.0	13	2.4	4.7
4	tù	-14, -12, -10, -8, -6	-9.0	19	0.9	4.8
5	lóng	-15, -13, -11, -9, -7	-11.6	17	2.1	2.1
6	shé	-20, -18, -16, -14, -12	-16.9	16	3.1	-3.1
7	mǎ	-19, -17, -15, -13, -11	-14.9	18	4.6	-1.1
8	yáng	-20, -18, -16, -14, -12	-14.6	15	2.4	-0.9
9	hóu	-13, -11, -9, -7, -5	-9.1	21	3.8	4.7
10	jī	-26, -24, -22, -20, -18	-25.6	9	5.1	-11.8
11	gǒu	-20, -18, -16, -14, -12	-14.5	20	1.9	-0.8
12	zhū	-21, -19, -17, -15, -13	-16.4	15	3.8	-2.6
Average			-13.8	16	2.9	0.0

RMSE, root mean squared error; SNR, signal-to-noise ratio; SRT, speech reception threshold.

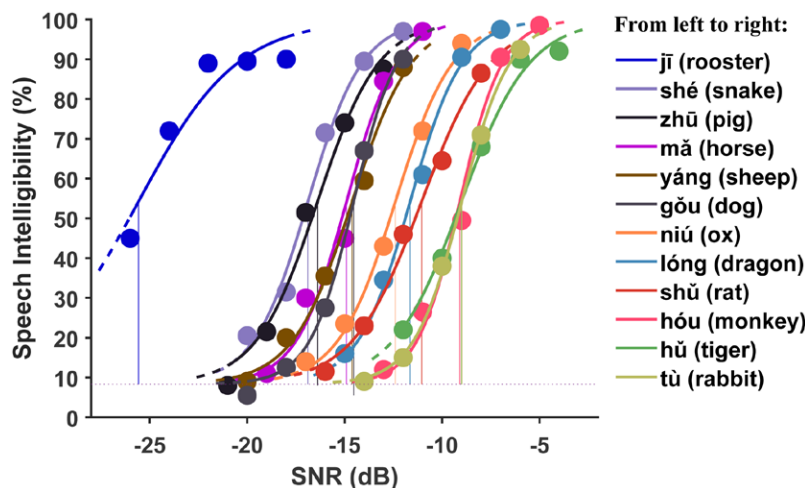


Fig. 3. Fitted psychometric functions of the 12 speech tokens. Each curve represents a speech token and the dots show the tested SNRs. The vertical thin lines indicate the SRTs. The horizontal dotted line shows the guess rate of $\gamma = 1/12$ in Eq. (1). SNR indicates signal-to-noise ratio; SRT, speech reception threshold

Downloaded from by BRNDMf5ePHkKH4TTImgenVA+ipjWIBvonh060Eigdm9T1vLQWJq+R2O4kRis8 on 02/26/2024

Initially, combinations of 3 distinct speech tokens selected from the set of 12 tokens were generated, resulting in a total of 220 combinations ($C_{12}^3 = 12 \times 11 \times 10 / (3 \times 2) = 220$). Furthermore, for each combination, a triplet was generated by randomizing the order of the 3 tokens within the combination and interspersing them with a silence of 150 ± 50 ms. Therefore, a set of 220 unique speech triplets (clean speech) was generated. This triplet set was balanced across words but was not balanced across triplet positions, that is, each token occurred 55 times in the 220 triplets but their positions were randomized. Subsequently, all 220 triplets were mixed with the masking noise at each of the 15 SNRs (-20 to $+8$ in 2-dB steps), resulting in 3300 audio files. A fixed noise level with varying speech levels was used for negative SNRs and a fixed speech level with varying noise levels was used for positive SNRs. This helped to avoid waveform clipping and provide a comfortable listening experience (Potgieter et al. 2016, 2018; De Sousa et al. 2018, 2020). To ensure noise refreshment, a noise fragment was randomly selected for each mixing. Although a certain degree of frozenness persists, where a given triplet presented at a specific SNR consistently retains the same frozen noise waveform, its effect on the freshness of the noise is relatively minor. This is because all triplets utilized for testing a particular user are inherently distinct, as detailed in the section The Adaptive Procedure. The noise duration was 600 ms longer than the triplet (300 ms at the start and end of the triplet).

Test Workflow • To ensure that the ZIN is easily self-administered by the largest possible population, the test workflow was designed with ease of self-administration in mind. Upon opening the test webpage, participants are asked to indicate their gender and age range after giving their consent. They are then instructed to wear headphones in a quiet environment and adjust the volume of the device to a comfortable level, while a test audio featuring a zodiac triplet in noise is played. A tutorial screen follows next, offering both textual and animated instructions on how to complete the test. Once the participant feels prepared, they may initiate the test by clicking a “Start” button, following a 3-second countdown. During the test, stimuli of triplets of zodiac animals in noise are presented diotically (same stimuli in both ears). No carrier phrase was presented before each triplet. Participants select their responses from a screen displaying 12 buttons (shown in Fig. 1), each including a cartoon image of a zodiac animal, the corresponding Chinese character, and its associated pinyin. The SRT is then measured using the adaptive procedure outlined below.

The Adaptive Procedure • To obtain reliable results with minimal time cost, an adaptive procedure was utilized for SRT measurement. This adaptive procedure was adapted from those previously reported in DIN studies by Smits et al. (2004) and Potgieter et al. (2016). As shown in Figure 1, the SRT is measured using a one-down one-up adaptive procedure consisting of 15 presentations. At the beginning of the procedure, a list of 15 distinct triplets was randomly chosen from the pool of 220 triplets described in the section Generation of Triplets. Each individual participant underwent testing with a unique list. However, precise control over the equal occurrence of each token at each position was not employed. The primary reason for employing 15 triplets instead of a lengthier list was to minimize the time required for the test. Given the test’s orientation toward the general public, we anticipated that a duration exceeding 3 minutes would significantly diminish the test’s acceptance and

participation rate. Moreover, a study with a test similar to the DIN test but using ecological sounds as target stimuli found that SRT convergence was reached at Trial 8 for the 1-up 1-down procedure (Denys et al. 2019). Another DIN study observed that the SRT-PTA correlation changed very little after 15 trials when using a step size of 2 dB (Watson et al. 2012). Therefore, we opted for the utilization of 15 triplets as a compromise between time efficiency and SRT convergence for our study. The procedure starts with an SNR of -2 dB for the first triplet. After the participant’s response, the SNR of the subsequent triplet is adjusted based on the accuracy of the previous response. If the response is incorrect, the SNR of the next triplet is increased by a step, whereas if the response is correct, the SNR of the next triplet is decreased by a step. The step size is set to 4 dB for the first five triplets and is then subsequently reduced to 2 dB for the remaining triplets. A response is considered correct when all zodiac animals are correctly identified in the correct order. The SRT is calculated as the average of the last six SNRs (i.e., SNRs of triplets 11–15, and the 16th SNR that is calculated but not presented).

Summary of development • The ZIN test used the 12 Chinese zodiac animals as speech materials and a quasi-stationary noise with the same long-term average spectrum as the masking noise. The speech tokens were carefully selected from natural recordings and were equalized in SNR for 50% intelligibility through an optimization process. The optimized speech tokens were combined into triplets, which were subsequently mixed with randomly selected noise segments to generate stimuli for the ZIN test.

Validations

The purpose of this section is to assess the effectiveness of the ZIN test in detecting hearing loss. A reliable hearing screening tool should be able to distinguish between those with normal hearing and those with hearing loss. To do that, a cutoff point for the ZIN test needs to be established, which would classify individuals as either having normal hearing or hearing loss. The ZIN test was then used to classify a group of listeners with various hearing abilities, and its effectiveness was assessed in comparison to pure-tone audiometry, the clinical gold standard.

In the work presented in this section, two groups of participants, with a total number of 256 individuals, were involved.

Normative Data • The normative group consisted of 140 individuals with normal hearing (73 females and 67 males, mostly aged between 16 and 30 years). The purpose of this group was to establish the normative data for the ZIN test, including the reference SRT, test-retest variability, time cost, and most importantly, a cutoff point for identifying hearing loss. Participants were recruited from university campuses and reported no history of ear disease or hearing issues. Participants were instructed to take the test on their own smartphones or tablets with headphones to mimic actual hearing screening scenarios. Among the 140 participants in this group, 40 individuals (25 females and 15 males, aged between 20 and 25 years) underwent two rounds of testing (test and retest), while the remaining 100 participants underwent a single round of testing. Test-retest variability was determined based on data from test and retest results from the above-mentioned 40 participants. Other measurements including the reference SRT, time cost, and cutoff point were determined based on the initial test of these 40 participants in conjunction with data collected from the other 100 participants

since individuals typically undergo the test only once in screening scenarios.

Validation Data • The validation group comprised 116 participants with various hearing abilities (53 females and 63 males, mean age 30.4 years ± 12.9 years [SD]). The purpose of this group was to examine the effectiveness of the ZIN test as a hearing screening tool. Sensitivity and specificity are the two metrics that are commonly used to evaluate the effectiveness of a hearing screening method. Sensitivity refers to the tool’s ability to accurately identify individuals with hearing loss, while specificity refers to its ability to accurately identify individuals without hearing loss. An effective hearing screening tool should strike a balance between sensitivity and specificity to ensure accurate and reliable results.

Participants in the validation group completed pure-tone audiometry to determine their true hearing condition, followed by testing with the ZIN test. Participants’ hearing levels were defined as the pure-tone average at octave frequencies from 0.5 to 4 kHz in their better ear, also known as the better-ear PTA. Using a cutoff better-ear PTA of 25 dB HL (although WHO has recommended a new cutoff point of 20 dB HL in 2021, the cutoff point of 25 dB HL was used in this study for better comparisons with previous studies (Bu et al. 2011; Hu et al. 2016; Gong et al. 2018; Qi et al. 2018)), the 116 participants included 90 with normal hearing and 26 with impaired hearing (audiograms shown in Fig. 4). Pure-tone audiometry was performed using a clinical audiometer in a sound booth following a standard clinical procedure. The ZIN test was conducted using participants’ own smartphones with commercial headphones.

RESULTS

Validations

Normative Data •

Reference Value • The SRT distribution of the normative group is presented in Figure 5A. The SRT values exhibit

a quasi-normal distribution with the highest density observed between -12 and -10 dB. The 50th percentile was at -11.0 dB (the median), the 75th percentile at -9.7 dB, and the 95th percentile at -7.7 dB. The SD was 1.6 dB. Based on these findings, the SRT value of -11.0 ± 1.6 dB, representing the 50th percentile and SD of the normative group, can be considered as the reference value for the ZIN test in individuals with normal hearing.

Time Cost • The inserted graph in Figure 5A shows how long it takes to complete the ZIN test. Most participants completed the test in between 1.5 and 3 minutes, with a median time cost was 2.4 min. The short time cost of the ZIN test makes it suitable for screening purposes, as a long test duration may reduce test uptake and completion.

Cutoff Point • Consistent with previous studies (Jansen et al. 2010; Smits et al. 2013; Potgieter et al. 2016), the cutoff point for the ZIN test was established using the upper 95th percentile point of the SRT distribution of the normal-hearing normative group. The rationale is that an SRT higher than the upper 95th percentile point can be considered significantly worse than that of normal hearing listeners at a confidence level of 95%. As a result, the cutoff point of -7.7 dB was used for the ZIN test as a standard to identify normal hearing and hearing loss.

Test-Retest Variability • The measurement error, that is, the test-retest variability, was defined as the SD of the within-subject differences between test and retest SRTs. In the 40 participants who underwent two rounds of testing, the calculated test-retest variability was 1.7 dB.

Validation Data •

Correlation With PTA • To evaluate the association between the ZIN test and pure-tone audiometry, the correlation between the ZIN SRT and the better-ear PTA was determined. As shown in Figure 5B, a significant correlation was found using a Pearson’s correlation analysis ($r = 0.82, p < 0.001$),

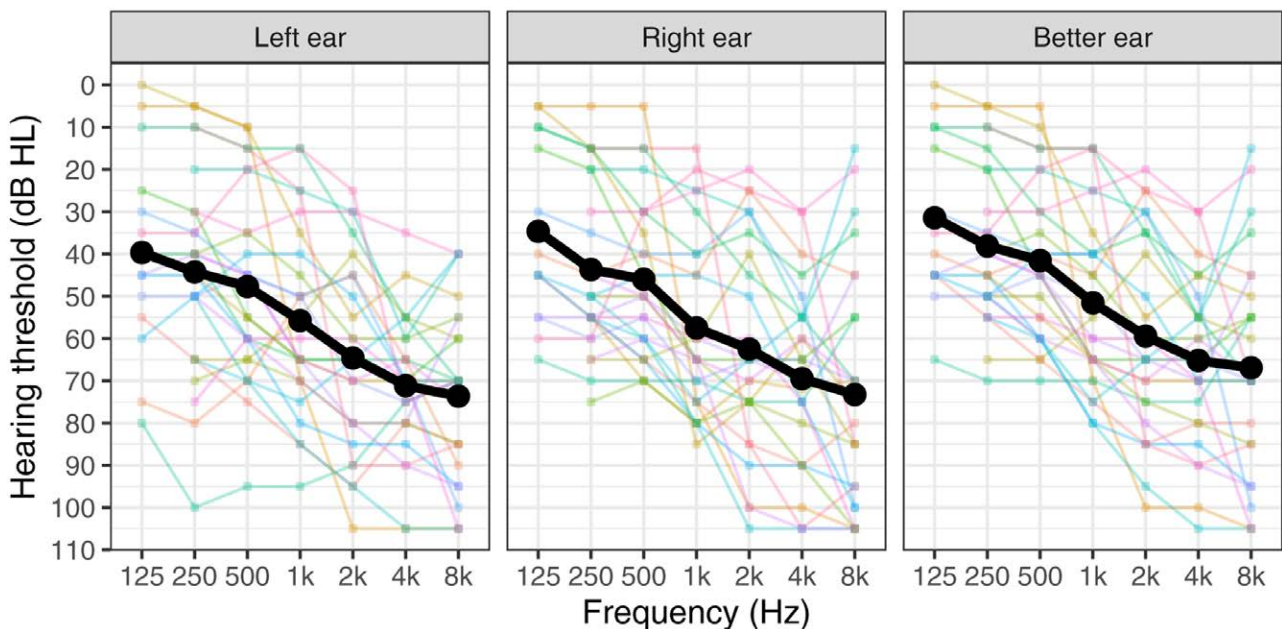


Fig. 4. Audiograms of listeners with hearing loss in the validation group (N = 26).

Downloaded from by BhDMf5ePHKbH4TTImqenVA+ipmWIBvonhoQ60Efgdm9T1vLQWJq+R2O4Kf58 on 02/26/2024

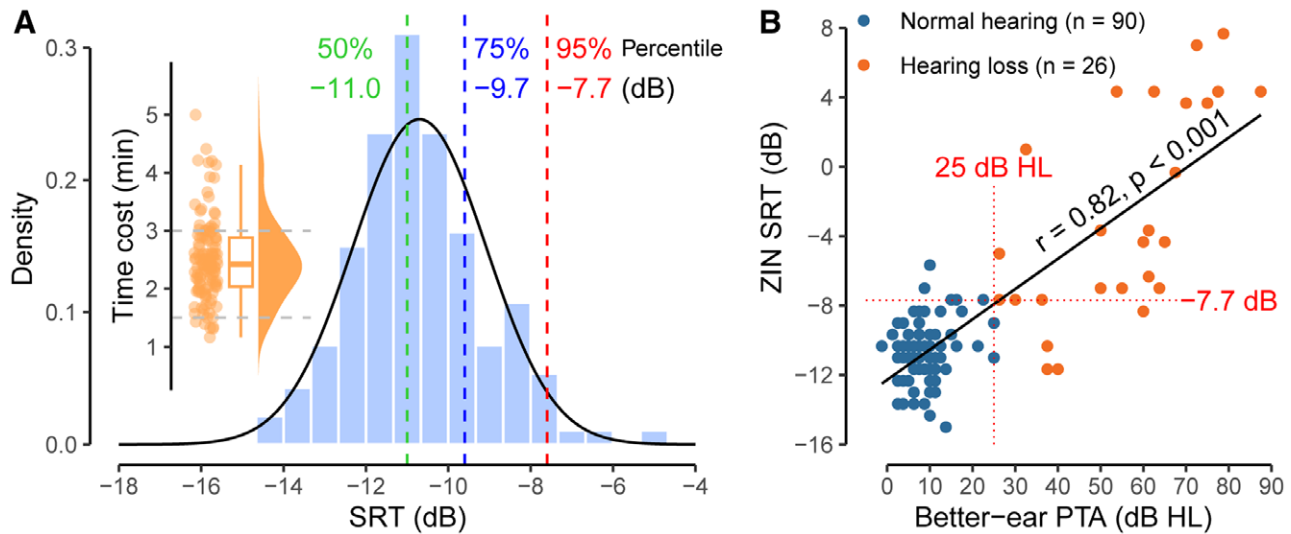


Fig. 5. Validation results. A, Normative data of the ZIN test obtained from the normative group ($N = 140$). The inserted graph shows the time cost of the ZIN test. B, Scatter plot of the ZIN SRT vs. the better-ear PTA from the validation group ($N = 116$). ZIN indicates Zodiac-in-noise.

indicating that the results of the ZIN test are associated with the better-ear PTA.

Sensitivity and Specificity • The sensitivity was calculated as the percentage of participants with hearing loss (PTA > 25 dB HL) who were correctly identified as having hearing loss by the ZIN test (SRT > -7.7 dB). The specificity was calculated as the percentage of participants with normal hearing (PTA ≤ 25 dB HL) who were correctly classified as having normal hearing by the ZIN test (SRT ≤ -7.7 dB). The results indicate that the ZIN test had a sensitivity of 0.85 and a specificity of 0.94. Further analysis indicates that the cutoff point of -7.7 dB was the optimal value that maximized the Youden Index (lowest misclassification) (Youden 1950) in the validation group.

Summary of validation • The developed ZIN test had a reference SRT of -11.0 ± 1.6 dB in normal-hearing listeners with a test-retest variability of 1.7 dB. Most participants could complete the test in 3 minutes or less. The result of the ZIN test was significantly correlated with the better-ear PTA ($r = 0.82$). The sensitivity and specificity for detecting a hearing loss of 25 dB HL or more at the better ear were found to be 0.85 and 0.94, respectively, using the selected cutoff point of -7.7 dB. These results demonstrated the ZIN test's effectiveness in detecting hearing loss.

Application in a remote hearing screening • A remote hearing screening was conducted using the ZIN test, starting from the World Hearing Day of 2022 (March 3, 2022). Data from the first week are shown in this section.

Study Sample • The publicity campaign through social networks enabled us to reach a large sample size of 30,552 participants in the first week. The participants' distribution by age and gender is presented in Figure 6A. The proportion of female participants (58%, $N = 17,594$) was higher than that of male participants (42%, $N = 12,958$). The three age groups with the largest number of participants for both genders were 16–20, 21–30, and 41–50 years. There were 2617 participants aged over 60 years and the number of participants decreased as

the age increased ($N = 1,842$ in 61–70 years, $N = 648$ in 71–80 years, $N = 78$ in 81–90 years, and $N = 49$ in 91–100 years).

SRT Distribution • Preliminary analysis revealed that 373 records reached the highest possible SRT of +8 dB. These records were excluded from further analysis as they were likely due to inattention. Also excluded was the age range 91–100, which had less than 50 observations. Data from the remaining 30,130 participants are presented in this section.

Figure 6B shows the distribution and cumulative distribution of SRTs of all 30,130 participants. The SRT distribution was positively skewed, with a long tail on the right side of the distribution. Using the cutoff point of -7.7 dB established in section Normative Data, participants whose SRT was higher than -7.7 dB (bins on the right side of the vertical dashed line in Fig. 6B) were classified as having hearing loss. A total of 6322 participants were identified as having hearing loss, resulting in an overall hearing loss prevalence of 21.0%.

We also looked into how age and gender affected ZIN SRT. The data in Figure 6B were divided into 10-year-wide age groups, grouped by gender, and summarized in Figure 6C. The cutoff border of -7.7 dB was also plotted. With aging, there was a general trend toward worse hearing (higher SRT). More than 75% of the participants had normal hearing in age groups up to 41–50, but this proportion dropped to 50% for the age group 61–70 and even to 25% or lower in age groups 71–80 and 81–90.

Females had better hearing than males in the age groups of 41–50 and older. Figure 6C shows that male participants had higher medians in SRT (worse hearing) than female participants in these age groups. To assess whether this gender difference was statistically significant, the nonparametric Mann–Whitney U test was conducted separately for each age group, given the skewed distribution of the data. The results indicate that the gender difference was statistically significant in the age group of 41–50 ($p < 0.001$), 51–60 ($p < 0.001$), 61–70 ($p < 0.001$), and 71–80 ($p = 0.003$). The lack of statistical significance in the age group of 81–90 may be attributed to the smaller number of participants in this age range,

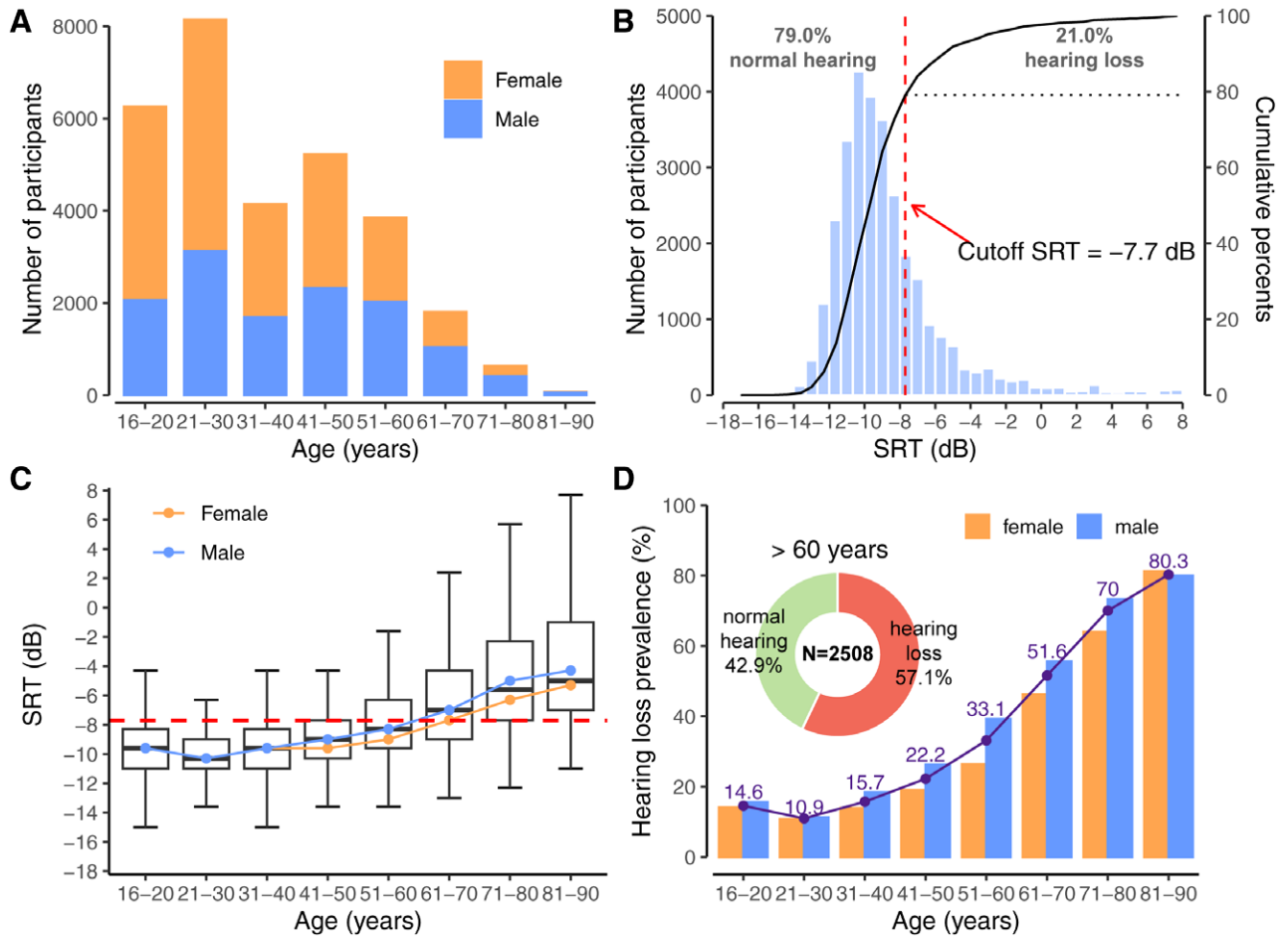


Fig. 6. Results of a large-scale hearing screening (N = 30,552). A, Histogram of participant age. B, SRT distribution and cumulative distribution across all the participants. Also shown are the cutoff SRT and the proportion of normal hearing and hearing loss based on SRTs. C, SRT distribution divided into 10-year age groups. Also shown are the lines of the medians by gender. The dashed line indicates the cutoff border of -7.7 dB. D, Hearing loss prevalence by age group. The line and numbers above the bars show the hearing loss prevalence averaged across gender. The inserted graph shows the ratio of hearing loss in participants aged over 60 years. SRT indicates speech reception threshold.

especially females (16 females and 55 males). The gender difference was not significant in the age groups 16–20 to 31–40 ($p > 0.05$).

Hearing deteriorated with age for both genders in age groups of 21–30 and older. As the gender difference was significant, a second analysis was performed to investigate how aging affected SRTs within each gender. The Mann–Whitney U test was conducted for adjacent age groups, that is, between one age group and its subsequent. To account for multiple comparisons, the significance level was set at $\alpha = 0.007$ as seven comparisons were made for each gender. The analysis revealed that the difference in SRT between all adjacent age groups was statistically significant ($p < 0.001$) for both genders.

Proportion of Participants Not Passing the Test • Finding the proportion of hearing loss in the study sample, or the percentage of participants who have hearing loss in the sample being studied is crucial in the context of hearing screening. In this study, we examined the proportion of participants not meeting the cutoff in our sample. The overall proportion of hearing loss in the entire study sample was 21.0%, as shown in Figure 6B.

Moreover, as discussed in the section SRT Distribution, our data showed that gender and age both had an impact on hearing

loss. As a result, we examined the proportion of hearing loss separately for each age group and gender. The findings are shown in Figure 6D, and they exhibited a similar trend to that of the SRT distribution graph in Figure 6C. Specifically, the proportion was lowest in the age group of 21–30 and increased with age, reaching 80.3% in the oldest age group of 81–90. The proportion in females was comparable to that of males in young age groups from 16–20 to 21–30 but was lower than that of males in age groups 31–40 and older. When pooling the age groups from 61–70 to 81–90, the mean proportion of hearing loss was 57.1% in the elderly study sample older than 60 years (see the graph inserted in Fig. 6D), highlighting the high proportion of hearing loss in this age range.

Summary • A remote hearing screening based on the ZIN test was conducted. The dissemination through social networks made it possible to collect data from 30,552 participants in just 1 week, which is a sizable sample size. Analysis of the data revealed that hearing deteriorated with age, and was gender-related, with females having better hearing than males in the age groups of 41–50 and above. The proportion of participants not meeting the cutoff was 20.1% overall in the study sample, with the elderly sample older than 60 years having a high proportion of 57.1%.

DISCUSSIONS

The main contribution of this work is the development of the ZIN test, a new Chinese speech-in-noise test for use in hearing screening in China. Moreover, this study conducted a remote and self-administered hearing screening using the developed ZIN test.

The ZIN Test

The ZIN test is a Chinese speech-in-noise test designed for hearing screening in China. It works by measuring an SRT using speech triplets in noise with an adaptive procedure. The 12 Chinese zodiac animals are used as speech materials to highlight the importance of consonant perception (Winn and Wright 2022). Its reference SRT is -11.0 ± 1.6 dB in normal-hearing listeners and can be completed in 3 minutes. The correlation between ZIN SRT and better-ear PTA was found to be 0.82. The test has a sensitivity of 0.85 and a specificity of 0.94 for detecting a 25 dB HL or more HL in the better ear at a cutoff value of -7.7 dB. The test-retest variability was 1.7 dB.

Comparisons With the DIN Test • Given that the ZIN test was developed and validated following a procedure similar to that of the DIN test, it is of interest to compare the results of the ZIN test with those of the DIN test, particularly the Chinese version of the DIN test. The main difference between the ZIN test and the Chinese DIN test lies in the speech materials, as well as the number of triplets used for testing and scoring.

The mean SRT of the ZIN test (-11.0 dB) is higher than the estimated diotic SRT of the Chinese DIN test. While Qi et al. (2018) did not explicitly report the mean SRT of the Chinese DIN test for normal-hearing (NH) listeners, it was approximated to be around -11 dB based on the concentration of data points between -12 and -10 dB, as illustrated in Figure 1 of their study. It is important to note that the data presented in their research pertained to monaural SRTs, which are estimated to be 1.4 dB higher (or worse) than diotic SRTs in NH listeners, as documented by Smits et al. (2006). Accordingly, the mean monaural SRT of -11.0 dB from the Qi et al. study translates to an estimated diotic SRT of approximately -12.4 dB, which is lower than the average diotic SRT of -11.0 dB in the ZIN test in this study.

The discrepancy in the average SRT between the two tests could potentially be attributed to the distinct speech materials employed in each study. Specifically, the Chinese DIN test utilized Chinese digits ranging from 0 to 9, whereas the ZIN test incorporated the 12 Chinese zodiac animals, which inherently possess a higher number of identical vowels. As a consequence, accurate perception of consonants becomes crucial to avoid confusion among speech items sharing identical vowels. Listeners may require higher SNRs to accurately perceive the consonants in the ZIN test, resulting in the relatively elevated diotic SRT compared to that of the Chinese DIN test. Other factors may also contribute to the difference in the average SRT, for example, the test environment and sample size. In the study of the Chinese DIN test (Qi et al. 2018), the average SRT was established from a small sample size (34 listeners with normal hearing) tested in a laboratory setting. In contrast, the average SRT in our study was based on a larger sample size (140 normal-hearing listeners) tested in participants' own environments, mimicking a true hearing screening.

The variability of the ZIN test, specifically the SD of SRTs in normal-hearing listeners (1.6 dB, as presented in the section Reference Value), was found to be larger when compared to the DIN test, including the Chinese DIN test. In most DIN studies, the variability of SRTs ranged within 1.0 dB, such as 1.0 dB for the South African English DIN (Potgieter et al., 2015), 0.9 dB for the Finnish DIN (Willberg et al. 2016), 0.5 dB for the French DIN (Jensen et al. 2010), and 0.62 to 0.72 for the Canadian DIN (Giguère et al. 2020). Although the SD of the Chinese DIN test was not explicitly reported, it was estimated to be less than 1.0 dB based on Figure 1 in Qi et al. (2018). This increased variability is closely related to the wider range between the average SRT and the cutoff. In our study, the range between the average SRT (-11.0 dB) and the cutoff (-7.7 dB) was 3.3 dB, which was wider than that observed in the Chinese DIN test (1.5 dB, Qi et al. 2018) and the South African DIN (2.2 dB, Potgieter et al. 2016).

Several factors may contribute to the larger variability observed in our study. One of the primary and possibly most significant factors is the increased measurement error resulting from a smaller number of triplet SNRs used for testing and scoring. In our study, we employed 15 triplets for testing, with the last 6 triplet SNRs used for calculating the SRT. In contrast, most DIN studies use 23–28 triplets for testing and 19–22 triplets for scoring. The rationale behind using fewer trials was to shorten the testing duration, thereby ensuring that it does not negatively impact the participation and completion rates. However, the findings of this study suggest that this was achieved at the cost of increased measurement error. The resulting measurement error, that is, the test-retest variability, was found to be 1.7 dB, which surpassed the measurement error of less than 1.0 dB observed in most DIN tests (refer to Table 1 in Van den Borre et al. 2021 for a review). Additionally, other factors such as the speech material, testing device, environment, and sample size may also contribute to the observed variability. In our study, the normative data were collected from a larger sample size, with participants using their own devices at their places. In contrast, participants in the DIN studies typically underwent testing in a smaller sample size using the same device in a controlled laboratory environment to establish normative data.

With a correlation coefficient of $r = 0.82$, the results of the ZIN test show a good association with PTA. This level of correlation is comparable to that of the Chinese and other language versions of the DIN test. The Chinese DIN test showed a correlation coefficient of 0.85 (Qi et al. 2018). Correlation coefficients for DIN tests in other languages have been found to range from 0.37 to 0.90 (see Van den Borre et al. 2021 for a review).

In terms of sensitivity and specificity, the ZIN test exhibited good performance for hearing screening purposes, with sensitivity and specificity of 0.85 and 0.94, respectively. Both values were higher than 0.80, indicating good accuracy. According to (Qi et al. 2018), the Chinese DIN test had a sensitivity/specificity of 0.94/0.94. Sensitivity/specificity values above 0.80 were observed for most DIN tests conducted in other languages, with some slightly lower values reported between 0.70–0.80 (see Van den Borre et al. 2021 for a review).

Comparisons With Sentence Speech-in-Noise Tests • The reference SRT of the ZIN test (-11.0 dB) is considerably lower than that of -4.3 dB for the Mandarin Hearing in Noise Test (MHINT, Wong et al. 2007), which is an open-set

sentence-in-noise test for Chinese listeners. Similar findings have been obtained in other languages, where the difference in SRTs obtained for triplets and sentence material presented in speech-shaped noise is evident. For instance, in Dutch, the reference SRT for a digit triplet-in-noise test is -11.2 dB (Smits et al. 2004), whereas it is -4.1 dB for a sentence-in-noise test (Versfeld et al. 2000). In British English, the reference SRTs for a digit triplet test is -12.1 dB, whereas it is -1.12 dB for a sentence test (Vainio et al. 2005; Zokoll et al. 2012). The lower SRTs for zodiac animal or digit triplets compared to meaningful sentences suggest that listeners perform better in simpler speech materials with fewer options (Jansen et al. 2012).

Another type of sentence test is the matrix test, which involves fixed-structure sentences composed of words from a limited word matrix. The matrix test can also be administered in a closed-set format, and SRTs obtained using matrix sentences are closer to those obtained using triplets compared to open-set sentence tests. The reference SRT obtained using the ZIN test (-11.0 dB) is closely aligned with that of the Mandarin Chinese matrix (CMNmatrix) sentence test (-11.2 dB) as reported by Hu et al. (2018). Similar results have been reported in other languages such as the Finnish DIN test (-10.8 dB) and the Finnish matrix test (-10.1 dB) (Dietz et al. 2014; Willberg et al. 2016).

Remote Hearing Screening in China

To the best of our knowledge, this study is the first large-scale remote hearing screening in China. With a substantial sample size of 30,552 adults, our study offers valuable insights into the proportions of hearing loss within this study cohort. Nonetheless, it is crucial to exercise caution when extrapolating the findings of this study to the broader Chinese population, as we did not employ any sampling controls to ensure the representativeness of the study sample. The overall proportion of hearing loss was 21.0% in the study sample. The proportion was noticeably higher in the study sample aged over 60 years, with 57.1% of participants experiencing hearing loss. The study also revealed that hearing deteriorated with age and that females had better hearing than males in the age groups of 41–50 and older. Our findings underscore the importance of early detection and prevention of hearing loss, especially among older adults and men. Therefore, this work contributes to increasing public awareness of hearing health in China.

The overall proportion of hearing loss based on the ZIN test (21.0%, see section Application in a Remote Hearing Screening) is consistent with previous research. For example, the Global Burden of Disease study conducted in 2019 estimated that 20.3% of the global population had varying degrees of hearing loss (GBD 2019 Hearing Loss Collaborators 2021). A recent study in a large sample ($N = 45,052$) based on better-ear PTAs found that the hearing loss prevalence in individuals older than 15 years was 19.2% (7374 out of 38,324), derived from Table 5 in Hu et al. (2016), which is also consistent with the proportion of 21.0% in the present study.

This study highlights the high proportion of hearing loss among people older than 60 years, with approximately 57.1% of participants experiencing hearing loss. This finding is consistent with previous studies that showed high rates of hearing loss in older adults. For example, Gong et al. (2018) reported data from hearing screening using pure-tone audiometry for 6984 older adults in China. The high proportion of hearing loss in

that study was 58.85%, which is comparable to that found in the present study. Similarly, an earlier study by Bu et al. (2011) reported a prevalence of hearing loss of 59.93% among older adults in China using methods similar to Gong et al. (2018). These consistent findings underline the importance of paying attention to the hearing health of older adults. Hearing loss is a common condition among this population and our study reinforces the need for increased awareness and resources to address this issue.

Our study also identified significant associations between hearing loss and age and gender. Older people and men were more likely to have hearing loss, which is consistent with previous studies in China (Hu et al. 2016; Gong et al. 2018) as well as in other countries (Cruickshanks et al. 1998; Helzner et al. 2005; Lee et al. 2015). Age-related hearing loss, or presbycusis, is a common condition among older adults, and our results highlight the need for increased attention to this issue among this population. The higher proportion of hearing loss among men is also a notable finding, which can be attributed in part to differences in occupational or environmental exposure to noise. Our study adds to the growing body of evidence supporting the importance of early detection and prevention of hearing loss, especially among older adults and men.

The consistency between the outcomes of the remote hearing screening based on the ZIN test and the in-person screenings based on pure-tone audiometry provides further evidence for the validity and efficiency of the ZIN test. The ZIN test is free of charge, self-administered, fast, and easy to interpret, making it an ideal tool for personal use or population-based screening. It can significantly reduce the financial and time cost of large-scale hearing screening. For example, a hearing screening with a comparable sample size to this study but using pure-tone audiometry was completed with an extremely high cost of having 144 primary investigators working in four provinces for over 1 year (Hu et al. 2016). On the contrary, the remote hearing screening in this study was completed in just 1 week with a negligible financial cost.

The COVID-19 pandemic has highlighted the need for remote healthcare services. An area where such services could be especially valuable is hearing screening. Remote hearing screening can significantly reduce the financial and time costs of conventional in-person hearing screening and could be especially beneficial for individuals who may be at risk of hearing loss but lack access to hospitals or clinics. The inclusion of a significant sample size of 30,552 participants in this study highlights the capacity of remote hearing screening to meet a critical public demand for hearing screening. Furthermore, the rapid recruitment of such a large study sample in a short period of 1 week suggests that social networks can play a positive role in facilitating the implementation of remote hearing screening and raising public awareness of hearing loss.

Limitations and Future Directions

One notable limitation of this study pertains to the sampling approach utilized during participant recruitment. No sampling control was applied to ensure the study sample was representative of the Chinese population in terms of socio-economic, lifestyle, and other factors that could affect hearing health. Our remote hearing screening was advertised on social media during the week of World Hearing Day without a control mechanism.

This approach may have introduced selection bias into the study sample. It is plausible that individuals who perceived themselves as having hearing issues and were eager to seek more information about their hearing health were more likely to participate, whereas those who believed their hearing was unaffected might be less motivated to engage in the study. Consequently, the study sample might not accurately reflect the diverse socio-demographic distribution of the broader Chinese population. This sampling limitation warrants caution when generalizing our findings beyond the context of this specific study sample.

An important future direction is to apply sampling controls to ensure that the study sample matches the socio-demographic distribution of the target population under study. Some in-person hearing screening programs have implemented sampling controls and provided reliable data on the hearing health of the target population. For example, Hu et al. (2016) is the most recent large-scale epidemiological study on hearing impairment targeted at the comprehensive Chinese population. In that study, probability proportionate to size sampling was used to ensure a representative study sample. The results obtained from that study hold significant importance in providing an accurate understanding of the current prevalence and influencing factors of hearing impairment in China. Remote hearing screening offers the advantage of efficient data collection by reducing both time and financial burdens. However, to ensure the validity and broader applicability of the findings obtained through remote hearing screenings, it is essential to implement appropriate sampling controls. Only with such controls in place, can we enhance the generalizability of the results and draw more reliable conclusions out of remote hearing screenings.

The current implementation of the ZIN test presents the same stimuli to both ears (i.e., diotic) and, therefore, is insensitive to asymmetrical or unilateral hearing loss due to the dominance of the better ear (Potgieter et al. 2018). Testing two ears consecutively (i.e., monaural) can identify asymmetrical or unilateral hearing loss, but this method doubles the test time cost, which may reduce test uptake and completion. Recently, researchers have explored a new test paradigm using antiphase stimuli, where the target speech is phase-inverted between the ears while the masking noise is kept in phase. This new test paradigm has been implemented in the DIN tests in South Africa and France, and there are also some pilot test data on a Chinese version of the antiphase DIN test in China (Fu et al. 2021; Gu et al. 2021). Studies have shown that this new paradigm markedly improves the sensitivity to detect asymmetrical hearing loss while minimizing the test duration by testing binaurally (De Sousa et al. 2020, 2022; Ceccato et al. 2021). Therefore, a future version of the ZIN test incorporating antiphase stimuli is worthwhile.

The study sample in this study is limited to adults, and the applicability of the ZIN test for children has not been investigated. Although China does not have a regular program for evaluating children's hearing, a suitable hearing test is urgently needed to screen children. Previous studies have extended the use of some hearing tests originally designed for adults to children, such as the DIN test (Denys et al. 2018; Koopmans et al. 2018; Vroegop et al. 2021; Wolmarans et al. 2021). Alternatively, other studies have developed new tests for hearing screening in children, utilizing ecological sounds (e.g., sounds of dogs, cats, birds, piano, etc.) (Denys et al. 2019), simple common words (Sheikh Rashid et al. 2017), and other similar materials.

All these tests use simple materials that have low linguistic requirements and obtain responses from children using simple images. The ZIN test has the potential to be an effective hearing screening tool for children because of its straightforward materials. The 12 zodiac animals should be familiar to children. However, further research is still needed to determine the feasibility and reliability of the ZIN test as a hearing screening tool for children.

SUMMARY AND CONCLUSION

The ZIN test, a hearing screening test based on speech-in-noise recognition, has been developed and validated as a reliable and efficient tool for hearing screening in China. The test adaptively measures SRT in speech-shaped noise using triplets of speech tokens that have been optimized for equalizing the SNR for 50% recognition. The use of the 12 Chinese zodiac animals as speech materials highlights the importance of consonant perception. The ZIN test has been found to have significant correlations with the better-ear hearing threshold ($r = 0.82$), with high sensitivity and specificity values (0.85 and 0.94, respectively) in detecting a hearing loss of 25 dB HL or more at the better ear.

A large-scale remote hearing screening was conducted using the ZIN test, with a sample size of 30,552 participants in just 1 week. The overall hearing loss proportion was 21.0% across the study sample and was high (57.1%) in the elderly sample aged above 60 years. Age and gender were found to be significantly associated with hearing loss, with older individuals and men being more likely to have hearing loss. These remote outcomes agree well with previous large-scale hearing screenings based on pure-tone audiometry completed with high financial and time costs.

In conclusion, the ZIN test has been demonstrated to be a valid and efficient tool for hearing screening in China. The application of the ZIN test in remote hearing screening has the potential to improve access to hearing healthcare for the general public.

ACKNOWLEDGMENTS

We would like to thank all participants who generously gave their time to participate in this study.

This work was supported by the National Natural Science Foundation of China (12374448), the Guangdong Basic and Applied Basic Research Foundation Grant (2022A1515011361), and the Shenzhen Fundamental Research Program (20220809191805001). The funding organization had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the decision to submit the article for publication; or in the preparation, review, or approval of the article.

H.L.Z. designed the experiments, conducted the experiments in the development stage, analyzed the data, and wrote the paper. X.H.L. and P.N.W. designed and conducted the validation experiment. S.D.S., X.W., and Y.Y.K. are employed by Tencent. S.D.S., W.X., and Y.Y.K. designed and conducted the remote hearing screening. J.W.L. and Y.M.W. designed and conducted the normative experiment. Q.L.M. and N.H.Z. conceived the test, designed the experiments, and discussed the results and implications. All authors commented on the article.

This study was approved by the Ethical Review Board of Shenzhen University.

The data that support the findings of this study are available upon reasonable request to the corresponding author.

The authors have no conflicts of interest to disclose.

Address for correspondence: Nengheng Zheng, Guangdong Key Laboratory of Intelligent Information Processing, College of Electronics and Information Engineering, Shenzhen University, Shenzhen, 518060, China. E-mail: nhzheng@szu.edu.cn

Received April 24, 2023; accepted September 23, 2023; published online ahead of print December 8, 2023.

REFERENCES

- Akeroyd, M. A., Arlinger, S., Bentler, R. A., Boothroyd, A., Dillier, N., Dreschler, W. A., Gagné, J. -P., Lutman, M., Wouters, J., Wong, L., et al. (2015). International Collegium of Rehabilitative Audiology (ICRA) recommendations for the construction of multilingual speech tests: ICRA working group on multilingual speech tests. *Int J Audiol*, 54(sup2), 17–22.
- Almufarrij, I., Dillon, H., Dawes, P., Moore, D. R., Yeung, W., Charalambous, A. -P., Thodi, C., Munro, K. J. (2022). Web- and app-based tools for remote hearing assessment: A scoping review. *Int J Audiol*, 1, 14.
- Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—A review. *Int J Audiol*, 42, 217–220.
- Bilger, R. C., & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. *J Speech Hear Res*, 19, 718–748.
- Bu, X., Liu, C., Xing, G., Zhou, L., Liang, C., Zheng, Y., Meng, J., Wang, Y., Yang, C., Liu, Y., Du, B., Zhang, B., Du, B. (2011). WHO ear and hearing disorders survey in four provinces in China. *Audiol Med*, 9, 141–146.
- Ceccato, J. -C., Duran, M. -J., Swanepoel, D. W., Smits, C., De Sousa, K. C., Gledhill, L., Venail, F., Puel, J. -L. (2021). French version of the antiphase digits-in-noise test for smartphone hearing screening. *Front Public Health*, 9, 725080.
- Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R., Mares-Perlman, J. A., Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin: The epidemiology of hearing loss study. *Am J Epidemiol*, 148, 879–886.
- Denys, S., Hofmann, M., Luts, H., Guérin, C., Keymeulen, A., Van Hoeck, K., van Wieringen, A., Hoppenbrouwers, K., Wouters, J. (2018). School-age hearing screening based on speech-in-noise perception using the digit triplet test. *Ear Hear*, 39, 1104–1115.
- Denys, S., De Laat, J., Dreschler, W., Hofmann, M., van Wieringen, A., Wouters, J. (2019). Language-independent hearing screening based on masked recognition of ecological sounds. *Trends Hear*, 23, 2331216519866566.
- De Sousa, K. C., Swanepoel, D. W., Moore, D. R., Smits, C. (2018). A smartphone national hearing test: Performance and characteristics of users. *Am J Audiol*, 27, 448–454.
- De Sousa, K. C., Swanepoel, D. W., Moore, D. R., Myburgh, H. C., Smits, C. (2020). Improving sensitivity of the digits-in-noise test using antiphase stimuli. *Ear Hear*, 41, 442–450.
- De Sousa, K. C., Smits, C., Moore, D. R., Myburgh, H. C., Swanepoel, D. W. (2022). Diotic and antiphase digits-in-noise testing as a hearing screening and triage tool to classify type of hearing loss. *Ear Hear*, 43, 1037–1048.
- Dietz, A., Buschermöhle, M., Aarnisalo, A. A., Vanhanen, A., Hyyrynen, T., Aaltonen, O., L’opp’onen, H., Zokoll, M. A., Kollmeier, B. (2014). The development and evaluation of the Finnish matrix sentence test for speech intelligibility assessment. *Acta Otolaryngol*, 134, 728–737.
- Fu, X., Li, Y., Gu, X. (2021). Application of hearing handicap inventory for adult screening and digits in noise in patients with hearing loss. *Chin Arch Otolaryngol Head Neck Surg*, 28, 340–343.
- GBD 2019 Hearing Loss Collaborators. (2021). Hearing loss prevalence and years lived with disability, 1990–2019: Findings from the global burden of disease study 2019. *Lancet*, 397, 996–1009.
- Giguère, C., Lagacé, J., Ellaham, N. N., Pichora-Fuller, M. K., Goy, H., Bégin, C., Alary, E., Bowman, R. (2020). Development of the Canadian digit triplet test in English and French. *J Acoust Soc Am*, 147, EL252–EL258.
- Gong, R., Hu, X., Gong, C., Long, M., Han, R., Zhou, L., Wang, F., Zheng, X. (2018). Hearing loss prevalence and risk factors among older adults in China. *Int J Audiol*, 57, 354–359.
- Gu, X., Li, Y., Fu, X. (2021). Application of digits in noise test in hearing screening. *Chin Arch Otolaryngol Head Neck Surg*, 28, 344–347.
- Guo, Z., Yu, G., Zhou, H., Wang, X., Lu, Y., Meng, Q. (2021). Utilizing true wireless stereo earbuds in automated pure-tone audiometry. *Trends Hear*, 25, 23312165211057367.
- Heine, C., Browning, C. J., Gong, C. H. (2019). Sensory loss in China: Prevalence, use of aids, and impacts on social participation. *Front Public Health*, 7, 5.
- Helzner, E. P., Cauley, J. A., Pratt, S. R., Wisniewski, S. R., Zmuda, J. M., Talbot, E. O., De Rekeneire, N., Harris, T. B., Rubin, S. M., Simonsick, E. M., Tykavsky, F. A., Newman, A. B. (2005). Race and sex differences in age-related hearing loss: The health, aging and body composition study. *J Am Geriatr Soc*, 53, 2119–2127.
- Hu, X., Zheng, X., Ma, F., et al. (2016). Prevalence of hearing disorders in China: A population-based survey in four provinces of China. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, 51, 2819–2824.
- Hu, H., Xi, X., Wong, L. L., Hochmuth, S., Warzybok, A., Kollmeier, B. (2018). Construction and evaluation of the Mandarin Chinese matrix (CMNmatrix) sentence test for the assessment of speech recognition in noise. *Int J Audiol*, 57, 838–850.
- ISO 8253-3 (2012). Acoustics. audiometric test methods-part 3: Speech audiometry.
- Jansen, S., Luts, H., Wagener, K. C., Frachet, B., Wouters, J. (2010). The French digit triplet test: A hearing screening tool for speech intelligibility in noise. *Int J Audiol*, 49, 378–387.
- Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., Fraysse, B., Vorm’es, E., Frachet, B., et al. (2012). Comparison of three types of French speech-in-noise tests: A multi-center study. *Int J Audiol*, 51, 164–173.
- Koopmans, W. J., Goverts, S. T., Smits, C. (2018). Speech recognition abilities in normal-hearing children 4 to 12 years of age in stationary and interrupted noise. *Ear Hear*, 39, 1091.
- Kramer, S. E., Kapteyn, T. S., Festen, J. M., Kramer, S. E. (1998). The self-reported handicapping effect of hearing disabilities. *Int J Audiol*, 37, 302–312.
- Lee, J. S., Choi, H. G., Jang, J. H., Sim, S., Hong, S. K., Lee, H. -J., Park, B., Kim, H. -J. (2015). Analysis of predisposing factors for hearing loss in adults. *J Korean Med Sci*, 30, 1175–1182.
- Lyzenga, J., & Smits, C. (2011). Effects of coarticulation, prosody, and noise freshness on the intelligibility of digit triplets in noise. *J Am Acad Audiol*, 22, 215–221.
- Nondahl, D. M., Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, R., Klein, B. E. K. (1998). Accuracy of self-reported hearing loss. *Int J Audiol*, 37, 295–301.
- Owens, E., Talbot, C. B., Schubert, E. D. (1968). Vowel discrimination of hearing-impaired listeners. *J Speech Hear Res*, 11, 648–655.
- Pienkowski, M. (2017). On the etiology of listening difficulties in noise despite clinically normal audiograms. *Ear Hear*, 38, 135–148.
- Potgieter, J. -M., Swanepoel, D. W., Myburgh, H. C., Hopper, T. C., Smits, C. (2016). Development and validation of a smartphone-based digits-in-noise hearing test in South African English. *Int J Audiol*, 55, 405–411.
- Potgieter, J. -M., Swanepoel, D. W., Myburgh, H. C., Smits, C. (2018). The South African English smartphone digits-in-noise hearing test: Effect of age, hearing loss, and speaking competence. *Ear Hear*, 39, 656–663.
- Qi, B., Zhang, T., Fu, X., Li, G. (2018). Establishment of the characterization of an adult digits-in-noise test based on internet. *Lin Chuang Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*, 32, 202–205.
- Sheikh Rashid, M., Dreschler, W. A., de Laat, J. A. (2017). Evaluation of an internet-based speech-in-noise screening test for school-age children. *Int J Audiol*, 56, 967–975.
- Smits, C., & Houtgast, T. (2005). Results from the Dutch speech-in-noise screening test by telephone. *Ear Hear*, 26, 89–95.
- Smits, C., Kapteyn, T. S., Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol*, 43, 15–28.
- Smits, C., Kramer, S. E., Houtgast, T. (2006). Speech reception thresholds in noise and self-reported hearing disability in a general adult population. *Ear Hear*, 27, 538–549.
- Smits, C., Theo Goverts, S., Festen, J. M. (2013). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. *J Acoust Soc Am*, 133, 1693–1706.
- Swanepoel, D. W., De Sousa, K. C., Smits, C., Moore, D. R. (2019). Mobile applications to detect hearing impairment: Opportunities and challenges. *Bull World Health Organ*, 97, 717–718.
- Vainio, M., Suni, A., Järveläinen, H., Järvikivi, J., Mattila, V. -V. (2005). Developing a speech intelligibility test based on measuring speech reception thresholds in noise for English and Finnish. *J Acoust Soc Am*, 118, 1742–1750.

- Van den Borre, E., Denys, S., van Wieringen, A., Wouters, J. (2021). The digit triplet test: A scoping review. *Int J Audiol*, *60*, 946–963.
- Versfeld, N. J., Daalder, L., Festen, J. M., Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *J Acoust Soc Am*, *107*, 1671–1684.
- Vlaming, M. S., MacKinnon, R. C., Jansen, M., Moore, D. R. (2014). Automated screening for high-frequency hearing loss. *Ear Hear*, *35*, 667–679.
- Vroegop, J., Rodenburg-Vlot, M., Goedegebure, A., Doorduyn, A., Homans, N., van der Schroeff, M. (2021). The feasibility and reliability of a digits-in-noise test in the clinical follow-up of children with mild to profound hearing loss. *Ear Hear*, *42*, 973–981.
- Wasmann, J. -W., Pragt, L., Eikelboom, R., Swanepoel, D. W. (2022). Digital approaches to automated and machine learning assessments of hearing: Scoping review. *J Med Internet Res*, *24*, e32581.
- Watson, C. S., Kidd, G. R., Miller, J. D., Smits, C., Humes, L. E. (2012). Telephone screening tests for functionally impaired hearing: Current use in seven countries and development of a US version. *J Am Acad Audiol*, *23*, 757–767.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Percept Psychophys*, *63*, 1293–1313.
- Willberg, T., Buschermöhle, M., Sivonen, V., Aarnisalo, A. A., Lööppönen, H., Kollmeier, B., Dietz, A. (2016). The development and evaluation of the Finnish digit triplet test. *Acta Otolaryngol*, *136*, 1035–1040.
- Winn, M. B., & Wright, R. A. (2022). Reconsidering commonly used stimuli in speech perception experiments. *J Acoust Soc Am*, *152*, 1394–1403.
- Wolmarans, J., De Sousa, K. C., Frisby, C., Mahomed-Asmail, F., Smits, C., Moore, D. R., Swanepoel, D. W. (2021). Speech recognition in noise using binaural diotic and antiphasic digits-in-noise in children: Maturation and self-test validity. *J Am Acad Audiol*, *32*, 315–323.
- Wong, L. L., Soli, S. D., Liu, S., Han, N., Huang, M. -W. (2007). Development of the Mandarin hearing in noise test (MHINT). *Ear Hear*, *28*, 70S–74S.
- World Health Organization (2021). World report on hearing.
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32–35.
- Zokoll, M. A., Wagener, K. C., Brand, T., Buschermöhle, M., Kollmeier, B. (2012). Internationally comparable screening tests for listening in noise in several European languages: The German digit triplet test as an optimization prototype. *Int J Audiol*, *51*, 697–707.
- Zokoll, M. A., Hochmuth, S., Warzybok, A., Wagener, K. C., Buschermöhle, M., Kollmeier, B. (2013). Speech-in-noise tests for multilingual hearing screening and diagnostics. *Am J Audiol*, *22*, 175–178.