Improving the Efficiency of the Digits-in-Noise Hearing Screening Test: A Comparison Between Four Different Test Procedures

[a] Tasneem Dambha
[a,b] De Wet Swanepoel
[a] Faheema Mahomed-Asmail
[a] Karina C. De Sousa
[c] Marien A. Graham
[d] Cas Smits

[a] Department of Speech-Language Pathology and Audiology, University of Pretoria, South Africa
[b] Ear Science Institute Australia, Subiaco, Western Australia
[c] Department of Science, Mathematics and Technology Education, Faculty of Education, University of Pretoria, South Africa
[d] Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Otolaryngology–Head and Neck Surgery, Ear and Hearing, Amsterdam Public Health Research Institute, the Netherlands

Abstract

Purpose: This study compared the test characteristics, test–retest reliability, and test efficiency of three novel digits-in-noise (DIN) test procedures to a conventional antiphasic 23-trial adaptive DIN (D23).

Method: One hundred twenty participants with an average age of 42 years (*SD* = 19) were included. Participants were tested and retested with four different DIN procedures. Three new DIN procedures were compared to the reference D23 version: (a) a self-selected DIN (DSS) to allow participants to indicate a subjective speech recognition threshold (SRT), (b) a combination of self-selected and adaptive eight-trial DIN (DC8) that utilized a self-selected signal-to-noise ratio (SNR) followed by an eight-trial adaptive DIN procedure, and (c) a fixed SNR DIN (DF) approach using a fixed SNR value for all presentations to produce a pass/fail test result.

Results: Test-retest reliability of the D23 procedure was better than that of the DSS and DC8 procedures. SRTs from DSS and DC8 were significantly higher than SRTs from D23. DSS was not accurate to discriminate between normal-hearing and hard of hearing listeners. The DF and DC8 procedures with an adapted cutoff showed good hearing screening test characteristics. All three novel DIN procedure durations were significantly shorter (< 70 s) than that of D23. DF showed a reduction of 46% in the number of presentations compared to D23 (from 23 presentations to an average of 12.5).

Conclusions: The DF and DC8 procedures had significantly lower test durations than the reference D23 and show potential to be more time-efficient screening tools to determine normal hearing or potential hearing loss. Further studies are needed to optimize the DC8 procedure. The reference D23 remains the most reliable and accurate DIN hearing screening test, but studies in which the potentially efficient new DIN procedures are compared to pure-tone thresholds are needed to validate these procedures.

Introduction

The World Health Organization (WHO) estimated the number of people with disabling hearing loss in 2018, and this estimate was 432 million adults. Furthermore, they projected that this number would rise to 900 million by 2050 (WHO, 2018a). Accessible screening to identify hearing loss in adults is an important priority to raise awareness of hearing loss; support earlier treatment; and prevent associated risks, such as cognitive decline (Koole et al., 2016; Livingston et al., 2020; Willberg et al., 2016). Self-assessed hearing screening tests using a speech-recognition-in-noise paradigm do not require calibrated and expensive equipment and can be easily accessed remotely. A popular screening test to determine the functional disability of hearing loss is the digits-in-noise (DIN) test (Jansen et al., 2010; D. R. Moore et al., 2014; Potgieter et al., 2016; Smits et al., 2004; Van den Borre et al., 2021; Watson et al., 2012; Zokoll et al., 2012).

The test uses spoken digit triplets (e.g., 3-4-8), mostly presented in speech-weighted masking noise. The first DIN was developed as the national telephone hearing test in the Netherlands (Smits et al., 2004). Due to high uptake of the original Dutch version, other countries developed the DIN test in different languages, such as French (Jansen et al., 2010), German (Zokoll et al., 2012), Finnish (Willberg et al., 2016), and American English (Watson et al., 2012).

Advancing technology has seen the DIN landline tele- phone screening test with limited bandwidth signal (300– 3400 Hz) move to broadband internet–based versions and mobile apps (Potgieter et al., 2016; Smits et al., 2004, 2006, 2013; Zokoll et al., 2012). Smartphone testing has the potential for widespread global access since 79% of the world's adult population was estimated to be smartphone owners by 2025 (Global System for Mobile Communications Association, 2019).

In 2016, the first national smartphone-based DIN test was launched as a free downloadable smartphone application (hearZA) in South Africa. Since its launch, more than 30,000 persons have downloaded the hearZA application to check their hearing status (De Sousa et al., 2020). The hearZA app monitors hearing health and provides users with a follow- up of their hearing status by promptly reminding the user to take the hearing test at least once a year (D. W. Swanepoel, 2017). Following the successful launch of hearZA using smartphone DIN testing, the WHO launched hearWHO in 2018, following the same testing procedure (D. W. Swanepoel et al., 2019; WHO, 2018b).

Most standardly used DIN procedures test each ear consecutively using an adaptive one-up, one-down procedure with a fixed step size of 2 dB to adjust the signal-to- noise ratio (SNR). The SNR decreases by 2 dB for a triplet identified correctly and increases by 2 dB when a triplet is entered incorrectly. The speech recognition threshold (SRT) is determined by averaging the SNRs of the presentations from the adaptive track, omitting the first few presentations, and representing the SNR where the listener can correctly recognize 50% of the triplets. When used for hearing screening, the SRT is compared to a predefined cutoff SNR to obtain a “pass” or “refer” result. Most of the current versions of the DIN are based on Smits et al. (2004, 2013) and use approximately 23 monaural or diotic presentations (Potgieter et al., 2016, 2018). The choice for this number of presentations is based on the desired test accuracy (i.e., the measurement error) and test duration. The measurement error or standard error of measurement (*SEM*) is inversely proportional to the square root of the number of presentations (Smits & Houtgast, 2006), but obviously, the test duration increases with the number of presentations. By ensuring high test–re- test reliability, hearing screening tests can efficiently identify even mild hearing losses, which is important for persons who regularly monitor their hearing using apps such as hearZA (D. W. Swanepoel et al., 2019). Short test time is desirable to reduce false-negative results caused by fatigue and to pro- mote better uptake of hearing screening tests in a consumer environment where rapid results are important (Denys et al., 2019; Potgieter et al., 2016; D. W. Swanepoel et al., 2019; Willberg et al., 2016). In this context, we call one test more *efficient* than another test if the test has the same measurement error as the other test but the result is achieved in a shorter time or when higher sensitivity and specificity are achieved in the same test time.

Several ways have been proposed to reduce the test time of the DIN without compromising test accuracy. First, a diotic test paradigm has been implemented to test both ears simultaneously instead of testing each ear sequentially (Potgieter et al., 2016, 2018; Smits et al., 2006). The diotic paradigm is time efficient because it reduces the test duration by approximately 50%. However, the test result may represent better ear performance and is not sensitive in detecting unilateral hearing loss (De Sousa et al., 2020). Therefore, second, De Sousa et al. (2020) introduced an antiphasic test paradigm. In this approach, identical broadband masking noise is presented to both ears, whereas the phase of the speech stimuli is inverted between ears (i.e., a 180° phase shift). The antiphasic test takes advantage of binaural unmasking (De Sousa et al., 2020) to improve the SRT by approximately 6–8 dB in normal-hearing individuals compared to the diotic approach (Smits et al., 2016). The antiphasic DIN is also sensitive to unilateral sensorineural and conductive hearing loss without extending test time through monaural assessment. De Sousa et al. found higher sensitivity (95%) and specificity (73%) for the antiphasic DIN than for the diotic DIN. A third method to optimize test efficiency was suggested by Denys et al. (2019). They employed a variable step size based on the number of correctly recognized digits within the triplet. Using this digit scoring with step sizes between 3 dB (none of the three digits correctly recognized) and −1 dB (all digits correctly recognized), they found SRTs that were not significantly different from the SRTs from the reference DIN. However, the measurement error for this procedure was significantly lower than that for the reference procedure with triplet scoring (Denys et al., 2019). They estimated that a

DIN with 17 presentations would be as accurate as a reference DIN procedure with 27 presentations for their procedure. Finally, a fourth method, a fixed SNR procedure, was proposed by Smits (2017). His approach was essentially different because he did not propose an adaptive procedure to estimate the SRT and then compare it to the test's cutoff value. In the procedure, all stimuli are presented at an SNR corresponding to the cutoff value (Smits, 2017). After each presentation, the probability of a pass or refer was estimated using Bayesian statistics, and the test is ended when the estimated probability of a pass or refer is higher than approximately 95%. Monte Carlo simulations showed that the DIN could theoretically be shortened to an average of approximately eight-digit triplet presentations (Smits, 2017). As far as we know, there have been no experimental data reported on the use of the fixed SNR procedure.

This research study aimed to explore the effect of different DIN procedures on its test characteristics and efficiency as a hearing screening test. The conventional anti- phasic DIN procedure was used as the reference test against which the new DIN procedures' accuracy, reliability, and performance were compared. The first new procedure explored the use of a subjective procedure to self- select the SNR using digit presentations. This method may require fewer presentations and a shorter test time. The second new procedure used a short adaptive approach to estimate the SRT, preceded by a self-selected SNR to approximate the starting SNR. Finally, the third procedure was the fixed SNR procedure (Smits, 2017).

Method

Institutional review board clearance was obtained prior to data collection commencement for the research study from the Humanities Research Ethics Committee, University of Pretoria (Protocol No. HUM011/1219).

Study Design and Participants

A cross-sectional, comparative study, including 120 participants, was performed. Participants' ages ranged be- tween 18 and 80 years, with an average age of 42 years (*SD* = 19), and all were proficient in English. Sixty-two percent of participants were women, and 38% were men. All participants provided consent and completed the ISO 989-1 checklist to ensure the ontologically normal candidacy criteria were met. The test–retest reliability, test characteristics, and efficiency of three different DIN procedures were compared to a conventional antiphasic 23-trial adaptive DIN (D23).

Procedures and Equipment

All participants completed eight DIN tests, which included a test and retest for each of the four different procedures. The four tests were presented in counterbalanced order using separate Latin square designs for both the test and retest sessions. A 30-min rest period between test and retest conditions was implemented to avoid participants fatiguing. The researcher visited the participant's home with the equipment (the smartphone device, headphones, sanitization tools, and paperwork). Testing took place in a quiet room at the participant's home, and low noise levels were ensured by closing windows that could allow outside noise to interrupt. All equipment was sanitized prior to and after testing the participant.

The initial study design was to include pure-tone audiometry. However, due to the COVID-19 pandemic and Level 5 lockdown regulations in South Africa, audiological soundproof booths and equipment could not be accessed to test participants as initially planned. The study used the national smartphone DIN test, hearZA, as the reference test to compare the three novel DIN procedures. The reference test has been validated and proven to be an accurate screening tool with high sensitivity and specificity and correlates strongly with pure-tone audiometry (r = .88; De Sousa et al., 2020).

A Huawei P30 lite smartphone run by Android Version 9.0 and connected to Sennheiser HD 280 Pro head- phones was used for the study. An Android application was installed to test the reference DIN (hearX Group), and an additional research Android application was installed to test the three novel DIN procedures.

The DIN test procedures in this study used digit triplets (e.g., 8-9-2), except for the self-selected DIN (DSS) procedure, which used random sequences of single digits, constructed from digits 0 to 9 spoken by a female speaker in South African English (Potgieter et al., 2016). For digit triplet presentations, the stimuli are as described by De Sousa et al. (2020). The noise starts 500 ms before the first digit and ends 500 ms after the last digit. The successive digits are presented with 200 ms of pause in be- tween, and a random jitter of 100 ms was applied to these pauses (Potgieter et al., 2016). The test implemented antiphasic stimuli by presenting the same-phased broadband speech-shaped white masking noise to both ears, while inverting the phase of the speech stimuli between the ears (i.e., 180° phase reversed; De Sousa et al., 2020).

Antiphasic D23

The D23 test was used as a reference test in this study. The test first required the participant to provide general information and read the on-screen instructions of the test, which was already described by the researcher before starting the test. Then, the participant was in- structed to select a comfortable presentation level of random consecutive digits by using the slider on the screen. The level increased (slide to the right) or decreased (slide to the left) when moving the slider. Next, the actual test started, and the participant was prompted to identify the three digits in the presence of competing noise (Potgieter et al., 2016). The participant was required to type on a keypad provided on-screen the three digits they recognized. If they were unsure, they had to guess. The test used fixed speech levels, adjusted the level of the masking noise when the SNR was positive, and used fixed noise levels for negative SNRs (De Sousa et al., 2020; Potgieter et al., 2016). The first digit triplet was presented at 0 dB SNR. The test followed a one-up, one-down adaptive procedure to estimate the SRT (De Sousa et al., 2020). The SNR of the initial three presentations was adjusted by either de- creasing by 4 dB for a correct response or increasing by 2 dB for an incorrect response. The remaining 20 SNRs were adapted in 2-dB steps. The last 19 SNRs were aver- aged to obtain the SRT (De Sousa et al., 2020; Potgieter et al., 2016). In total, there were 23 triplet presentations (De Sousa et al., 2020; Potgieter et al., 2016). Appendix Figure A1 provides screenshots of the D23 test described above.

DSS

The DSS used a continuous sequence of random consecutive antiphasic digits (0–9) presented in the presence of background noise (e.g., 1–4–7–9–5–3). Note that, unlike the other procedures, single digits were used. The digits were presented with 500-ms intervals between digits, and the masking noise was presented continuously. Five buttons were presented on the smartphone screen, and the participant was instructed with the text: “Adjust the volume (using the buttons below) until you can only hear the digits.” A verbal instruction was provided to the participants to use the buttons to adjust the volume of the digits in background noise until they could only just barely recognize the digits. Each button changed the SNR, ranging from the highest SNR (easiest) on top to the lowest SNR (most difficult) at the bottom. Once the participant made their selection, they were required to press “Next” to proceed to the following adjustment trial. Four adjustment trials were used, and each adjustment trial used a different adjustment sensitivity.

(a) The first adjustment trial used 4-dB SNR decrements between 0 dB SNR (top button) and −16 dB SNR (bottom button)

(b) The second adjustment trial used 2-dB decrements, with the top button corresponding to an SNR 2 dB higher than the selected SNR in the first adjustment trial.

(c) The third and fourth adjustment trials also used 2-dB decrements, with the top button corresponding to an SNR 2 dB higher than the selected SNR in the second adjustment trial. Thus, the fourth adjustment trial was similar to the third. For each adjustment trial, the presentations start with the top button selected. Appendix Figure A2 shows a screen- shot of the adjustment trial of the DSS test. The third adjustment trial SRT value was used in the data analysis.

Combination of Self-Selected and Adaptive Eight-Trial DIN

The combination of self-selected and adaptive eight- trial DIN (DC8) procedure had two phases. The first phase followed the same procedure as the DSS to allow the participant to self-select the SNR. However, just one adjustment trial with five buttons was used to bring the presentation level to roughly approximate the SRT quickly. The adjustment trial presented in the first phase used 4-dB SNR decrements between −4 dB SNR (top button) and −20 dB SNR (bottom button). Phase 2 of the DC8 procedure was a short antiphasic adaptive DIN with eight presentations, which started at the self-selected SNR. In this phase, the DC8 procedure presented digit triplets, which followed a similar testing procedure as the D23 procedure. For the first three presentations, the SNR decreased by 4 dB for a correct response and increased by 2 dB for an incorrect response. The remaining five presentations followed a 2-dB step size, and the final SRT was based on an average of these last five presentations. Appendix Figure A3 is a dis- play of Phases 1 and 2 of the DC8 test.

Fixed SNR DIN

As proposed by Smits (2017), this test procedure used a fixed SNR, corresponding to the cutoff value of −14.5 dB SNR established by De Sousa et al. (2020). The participant was instructed to select a comfortable presentation level of random consecutive digits by using a slider to increase (slide to the right) or decrease (slide to the left) the SNR of the digit triplets and choose “Next.” Then, two trial presentations were presented at −6.0 and −10.0 dB SNR to get used to the test. These presentations were not used in determining the test result. Then, the actual test started. The minimum number of presentations was set at 6, and the total number of presentations was variable, with a minimum of 8, when including the two trial presentations. After each presentation, the proportion of correct responses was calculated, and the probability that the true SRT was better than the cutoff value (pass) or worse than the cutoff value (refer) was estimated using Bayesian statistics. When the estimated probability for a pass or refer was higher than 95% or the maximum number of presentations (i.e., 25) was reached, the test stopped. Otherwise, another stimulus was presented, and the calculations were repeated. For example, if seven or more of the responses were correct after nine presentations, the DIN test ended with a pass; if two or less of the responses were correct, the DIN test ended with a refer. In all the other cases, another presentation followed. Appendix Figure A4 shows the test setup of the screenshots from the fixed SNR DIN (DF) test.

Statistical Analysis

All data were analyzed using SPSS (Version 26.0; IBM). Because the data were not normally distributed, even after transformation, nonparametric tests were used on the unaltered data. Nonparametric Spearman correlation coefficients were calculated between the SNRs obtained for the four different adjustment trials of the DSS method and D23. The strongest correlations with D23 were found for the third and fourth adjustment trials (nearly identical values); therefore, the SNRs from the third level were used in the analyses. The nonparametric Wilcoxon signed-ranks test was used to test for significant differences between mean SRTs for each test method. The nonparametric Spearman correlation was used to compute correlations between variables and to determine the test–retest reliability for quantitative variables. The test-retest reliability for qualitative variables was tested using Cohen's kappa. The *SEM* was determined from the differences between the test and retest SRT estimates obtained from each participant. Logistic regression was used to create pass rate functions for different DIN procedures, and sensitivity and specificity were also calculated for the different DIN procedures. The sensitivity of each new DIN test procedure was calculated by dividing the number of correctly identified participants with hearing loss by each procedure by the total number of participants with hearing loss. Likewise, the specificity was calculated by dividing the number of correctly identified participants without hearing loss by the total number of participants without hearing loss. Hearing loss was defined, using the reference DIN SRT (D23), as an SRT greater than -14.5 dB SNR (De Sousa et al., 2020) - due to the absence of pure-tone audiometric thresholds.

Results

Test-Retest Reliability

The scatter plots in Figure 1 show test versus retest data for the reference DIN (D23), the DSS, and the combi- nation DIN (DC8). Test-retest differences, represented by scatter around the line of equality, are generally smaller for D23 and largest for DSS. Table 1 shows the mean SRTs (and standard deviations) for test-retest conditions and other characteristics for the three novel DIN procedures. It shows that D23 has the strongest correlation between test and re-test SRTs and the smallest *SEM*. The DC8 test procedure shows a significant improvement of retest SRTs of 0.6 dB, indicative of a learning effect. Test-retest reliability of the fixed SNR procedure (DF), assessed by Cohen's kappa, is $\kappa = .630$, $p < .001$, indicating a substantial agreement (McHugh, 2012).

Relationships Between Results from the New DIN Procedures and the Reference DIN Procedure

The average of test and retest SRTs was calculated and used to explore the different procedures' relationships. Figure 2 shows the mean SRT from the DSS against D23 and the mean SRT from the DC8 against D23. Clearly, the self-selected SRTs are generally higher than the D23 SRTs because almost all data points lie above the line of equality (see Figure 2, left panel).

The correlation between DSS SRT and D23 SRT, $r_s(120) = .203$, $p < .001$, is weaker than the correlation be- tween DC8 SRT and D23 SRT, $r_s(120) = .696$, $p < .001$. The DSS SRTs are significantly higher than the D23 SRTs (mean difference of 5.8 dB, $z = 8.539$, $p < .001$), and the DC8 SRTs are also significantly higher than the D23 SRTs (mean difference of 0.99 dB, $z = 5.134$, $p < .001$). Thus, different cutoff values are needed when using the DSS, DC8, or D23 procedures for hearing screening.

Screening Characteristics of the Different DIN Procedures

The reference D23 has previously shown high sensitivity and specificity to detect hearing loss and been strongly correlated with pure-tone average (PTA) thresholds ($r = .88$; De Sousa et al., 2020). The cutoff value of the reference D23 is −14.5 dB SNR. Thus, ideally, each test should dis- criminate between participants with SRTs above and below this value. Therefore, we used this standard cutoff value to evaluate screening characteristics. To take into account the systematic difference of 5.8 dB between DSS SRT and D23 SRT and 0.99 dB between DC8 SRT and D23 SRT, we also adapted cutoff values of −8.7 dB SNR (−14.5 + 5.8) and −13.5 dB SNR (−14.5 + 0.99) for the DSS and DC8 procedures, respectively. Figure 3 shows the proportion of the tests resulting in "pass," as a function of the mean SRT of the D23 procedure. The solid lines represent the pass rate functions, which are logistic functions fitted to the data. The mean D23 SRT was rounded, and results were grouped according to the round SRT. Note that test and retest SRTs for the DSS and DC8 procedures were not averaged but treated as in- dependent measures. For each procedure, the raw data were fitted with a logistic function through a maximum likelihood procedure represented by solid lines in Figure 3.

Table 2 provides details of the pass rate functions and sensitivity and specificity of different DIN procedures. The slope of the pass rate function indicates the width of the range of SRTs where the pass rate drops from high to low values. The pass rate function of the DF procedure has the steepest slope and demonstrates the smallest region of SRTs where the accuracy of the test in discriminating between pass and fail is poor. The test characteristics of the DSS procedure are poor. Using an adjusted cutoff value improves the balance between sensitivity and specificity, but the resulting values are near chance level (= 0.50) for this procedure. Using an adjusted cutoff value for the DC8 procedures improves the specificity of the test but, of course, at the cost of lower sensitivity.

Figure 1.
Bivariate plots of the retest speech recognition threshold (SRT) against the initial test SRT for three different digits-in-noise (DIN) procedures. A small amount of jitter was added to each data point in the middle panel to avoid overlap. D23 = 23-trial adaptive DIN; DSS = self-selected DIN; DC8 = combination of self-selected and adaptive eight-trial DIN.

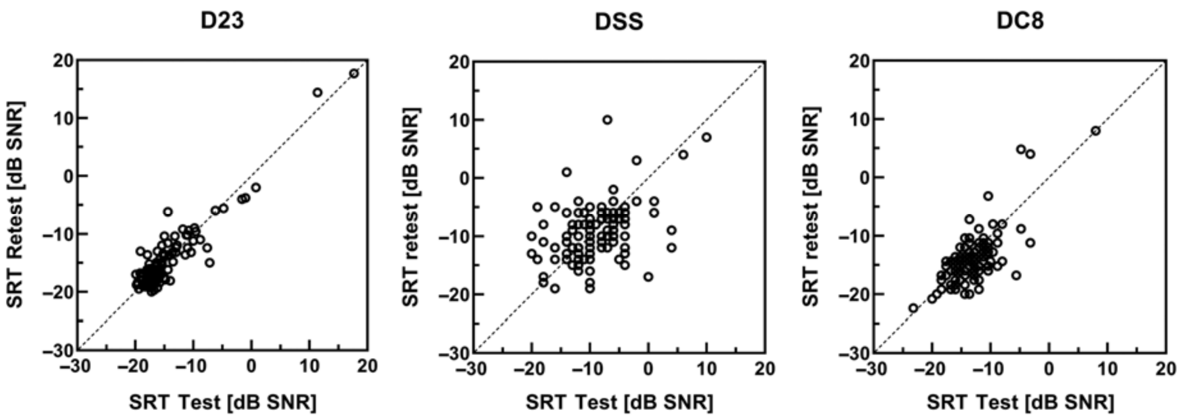


Table 1.
Mean speech recognition thresholds (SRTs) and standard deviations (*SD*s) for test–retest conditions and other characteristics across three digits-in-noise (DIN) procedures.

	Mean SRT (*SD*)			Spearman correlation	*SEM* (dB)
DIN test	Test	Retest	Difference (sig.)[a]	(sig.)	
D23	−14.8 (5.4)	−14.8 (5.4)	$z = -0.006$ ($p = .995$)	0.748 ($p < .001$)	1.5
DSS	−9.0 (5.2)	−9.1 (4.8)	$z = -0.320$ ($p = .749$)	0.409 ($p < .001$)	3.7
DC8	−13.5 (3.9)	−14.1 (4.4)	$z = -2.103$ ($p = .035$)	0.536 ($p < .001$)	2.2

Note. SRT values measured in dB SNR. D23 = standard adaptive 23-trial DIN; DSS = self-selected SRT; DC8 = combination of self-selected and adaptive eight-trial DIN.
[a]Difference between test and retest measured using the Wilcoxon signed-ranks test.

Test Time and Number of Presentations of the Different DIN Procedures

Figure 4 shows violin plots of the test durations of the different procedures. The average test duration of the reference D23 procedure was 2 min 16 s. The three novel DIN procedures all have significantly shorter test dura- tions than the reference D23 test, with average test dura- tions of 68, 57, and 47 s for the DF ($z = -7.744$, $p < .001$), DSS ($z = -9.150$, $p < .001$), and DC8 ($z = -9.192$, $p < .001$) procedures, respectively.

The DF procedure varies in the total number of presentations, with an average of 12.5 presentations, including the two trial presentations (*SD* = 5.6, range: 8– 25) to produce a pass/fail result. Fifty-one percent of the tests were completed after a minimum of eight presentations. The bubble plot in Figure 5 shows the number of presentations per test as a function of the average D23, with the size of the bubbles corresponding to the number of tests. The shape of the calculated distribution is as expected, with the highest number of presentations for participants with SRTs near the cutoff SNR (represented by the vertical dashed line; Smits, 2017).

Figure 2.
Bivariate plots of mean speech recognition thresholds (SRTs) from new digits-in-noise (DIN) procedures against mean SRTs from the reference DIN procedure. D23 = 23-trial adaptive DIN; DSS = self-selected DIN; DC8 = combination of self-selected and adaptive eight-trial DIN.

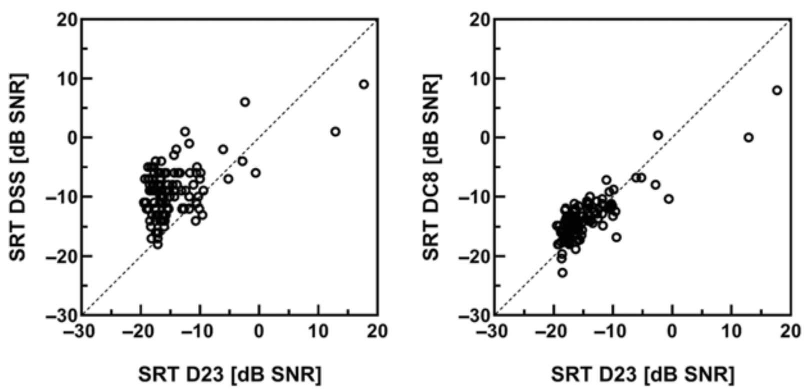
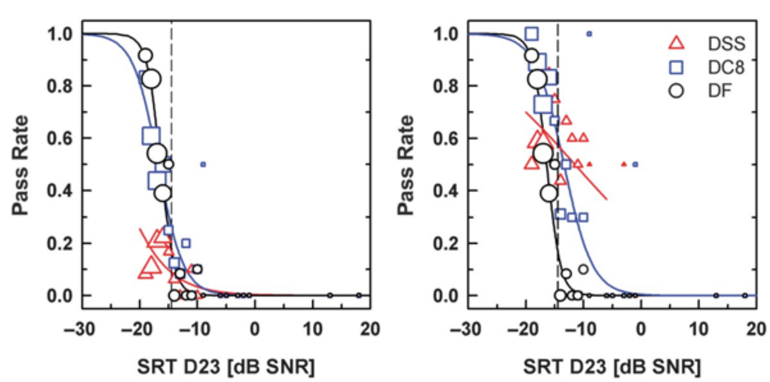


Figure 3.
Proportion of the new digits-in-noise (DIN) procedures, which result in “pass” as a function of the mean speech recognition threshold (SRT) of the D23 procedure. The left panel shows the results when using the cutoff value of −14.5 dB for all procedures, and the right panel shows the results when using the adapted cutoff values. D23 = 23-trial adaptive DIN; DSS = self-selected DIN; DC8 = combination of self-selected and adaptive eight-trial DIN; DF = fixed SNR DIN.



Discussion

The main aim of this study was to compare the test–retest reliability, test characteristics, and efficiency of three DIN procedures to a reference DIN test. Overall, all three novel DIN procedures had significantly shorter test durations of less than 1 min, which was better than the average duration of more than 2 min of the reference D23 test.

The conventional D23 method had the lowest *SEM* (1.5 dB SNR) calculated from the test–retest differences, indicating the highest test–retest reliability. The *SEM* is very similar to the reported *SEM* for the same test in the study by De Sousa et al. (2020).

DSS

The DSS procedure was based on a completely subjective measurement of how an individual perceives speech in the presence of background noise. Although the test is quick, it is not useful as a screening test because of the poor test characteristics and poor correlation with the reference test. Differences between test and retest SRTs are large, yielding an *SEM* of 3.7 dB SNR (see Table 1 and Figure 1).

Table 2.
An overview of the test characteristics of the pass rate functions and sensitivity and specificity of different digits-in-noise (DIN) test procedures.

Test procedure	Slope of pass rate function (dB^{-1})	Correctly classified (%)	Sensitivity (1 – FNR)	Specificity (1 – FPR)
Standard cutoff = −14.5 dB SNR				
DF	-0.21	74	0.97	0.64
DSS	-0.06	41	0.97	0.16
DC8	-0.11	65	0.91	0.54
Adjusted cutoff values				
DSS	-0.02	61	0.54	0.64
DC8	-0.10	80	0.73	0.83

Note. FNR = false-negative rate; FPR = false-positive rate; SNR = signal-to-noise ratio; DF = fixed SNR DIN; DSS = self-selected DIN; DC8 = combination of self-selected and adaptive eight-trial DIN.

Even more clear are Figure 3 and Table 2, which show that the test performs almost at chance level. Changing the cutoff value does not improve its performance as a screening test. The self-selected SNRs are much higher than the SRTs from the D23, demonstrating that participants choose very favorable SNRs where recognition probabilities are high. A similar test that allows the individual to self-select their perceived SNR is the Performance-Perceptual Test, which was developed and validated as a tool to measure discrepancies between objective and subjective measures of speech recognition in noise. The test compares the "perceived" SRT and the "measured" SRT (Saunders & Cienkowski, 2002; Saunders et al., 2004). In line with the study by Saunders et al. (2004), we found a significantly higher (z =

−8.264, $p < .001$) mean subjective (DSS, $M = -9.0$ dB SNR) compared to measured SRT (D23, $M = -14.8$ dB SNR). An important issue with the DSS procedure is the interpretation of the instruction for the participants. The instruction on the screen was different from the verbal instruction, and it is questionable if one of these instructions makes clear that the target setting refers to an SNR where one can just recognize the digits (or maybe recognize 50% of the triplets). Although a less ambiguous instruction could help to re- duce variability between participants or the difference be- tween DSS SRT and D23 SRT, it probably does not im- prove the measurement error.

Figure 4.
Violin plot showing the test duration of the different digits-in-noise (DIN) procedures. Horizontal solid lines depict median values. D23 = 23-trial adaptive DIN; DF = fixed SNR DIN; DSS = self-selected DIN; DC8 = combination of self-selected and adaptive eight-trial DIN.

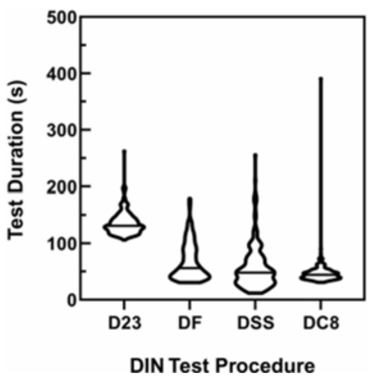
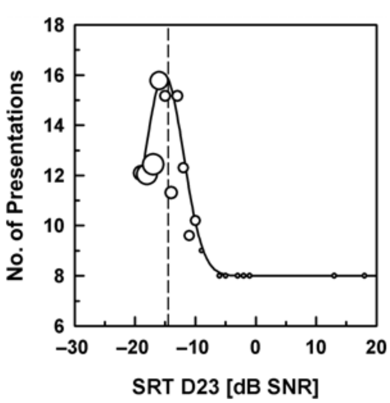


Figure 5.
Bubble plot showing the number of presentations for the fixed SNR digits-in-noise (DIN) procedure. The sizes of the bubbles correspond to the number of tests. The solid line serves as a guide to the eye. SRT = speech recognition threshold; D23 = 23-trial adaptive DIN.



DC8

Our results suggest that a combination test proce- dure could potentially be used for an efficient hearing screening. However, the current implementation with eight presentations following the self-selected SNR has a larger *SEM* (2.2 dB) and weaker correlation between test and re- test SRTs than the reference D23 (*SEM* = 1.5 dB). Second, the average DC8 SRTs are significantly higher than the D23 SRTs. Given the linear relationship between *SEM* and the inverse square root of the number of presentations, it can be estimated that approximately 17 presentations would be needed for the combination test procedure (DC8) to reach the *SEM* of 1.5 dB SNR from the D23 procedure. Obviously, the implication of increasing the number of presentations of the DC8 would be an increase in the test duration. A possible reason for the higher average DC8 SRTs than the average D23 SRTs is that some participants self-selected a relatively high SNR. Then, the first presentation of the adaptive procedure was too far above the SRT, which causes a bias in the estimated SRT (Smits & Houtgast, 2006).

DF

The test characteristics for the DF procedure can be seen as the second-best choice as it showed high sensitivity, a shorter test time, and a much lower number of presentations (see Figure 3 and Table 2). The data are an experimental confirmation of the Smits (2017) simulations and demonstrate that the number of presentations can be reduced significantly when the aim of the screening test is solely to discriminate between normal-hearing and hard of hearing participants. A disadvantage of the procedure is that it does not provide direct information about the severity of the hearing loss as the other test procedures do. The DF's average number of presentations was 12.5, which showed a 45.7% decrease compared to the D23. The average number of presentations is higher than the value of 8.3 from the Smits simulations, but the mini- mum number of presentations in the present implementation of the procedure was set at 8 and included two trial presentations. When taking these differences into account, the estimated number of presentations in our experimental study and the estimated number of presentations in the simulations from Smits are very similar. Of course, the average number of presentations also depends on the distribution of the SRTs.

Study Limitations

One limitation of this study is the lack of pure-tone audiometry thresholds. We used the average of two DIN SRTs (D23) as a reference. This can be considered a valid method to compare the quality of the different test procedures to differentiate participants with SRT above or be- low a cutoff SNR. However, because the reference SRT is not error free, some of the true test characteristics may be somewhat better than reported here (e.g., the slope of the pass rate function).

Ideally, pure-tone audiometry is used as a reference measure when evaluating DIN tests. The reference DIN test, D23, as used in this study has previously been vali- dated as an accurate and reliable screening tool (De Sousa et al., 2020). It shows a high correlation with PTA (r = .88) and good sensitivity and specificity in detecting hearing loss. Furthermore, a speech-in-noise test allows for a better understanding of a person's hearing ability com- pared to pure-tone audiometry, as it represents hearing in everyday life. Other research studies have supported the use of speech-in-noise tests due to its ability to better depict hearing loss over pure-tone audiometry (Bergman, 1971; Vermiglio et al., 2012). Other DIN tests have also shown good correlations with PTA and, thus, can be justified as an acceptable stand-in test for pure-tone audiometry (Jansen et al., 2010; Potgieter et al., 2016; Smits et al., 2004; Van den Borre et al., 2021). However, when new and potentially efficient DIN procedures are considered for implementation, a study is needed in which the DIN procedures are compared to pure-tone thresholds to validate these procedures.

The end goal of smartphone-based DIN testing is to have a self-administered, easy-to-comprehend screening test. In line with our study, it was noted in the initial uptake of the South African DIN test (hearZA) that the median age was 37 years (K. C. Swanepoel, 2018, p. 51). Overall, more younger persons downloaded the initial hearZA app than older adults (K. C. Swanepoel, 2018, p. 51). A rea- son for this and for future teleaudiology practice is the concern that older adults do not have strong computer literacy skills, which may reflect overall limited uptake of health care technologies (A. N. Moore et al., 2015; K. C. Swanepoel, 2018, p. 51). For this study, the researcher was present to give the participants clear instructions on navigating the test. However, the complexity of having two different test phases in the DC8 may compromise the reliability when testing children and older adults who do not readily understand the test procedure. The interpretation of the test instructions for the DSS may differ between listeners, as some may choose the SNR at which they can clearly understand the digits instead of "barely understanding" the digits presented in background noise.

Conclusions

Overall, the reference DIN procedure (D23) pro- vides the most reliable and accurate test result. When implemented as a hearing screening tool, it can be easily accessed using a smartphone device and, on average, re- quires just over 2 min to complete. The DF is highly efficient and can be easily implemented because the cutoff SNR from the reference DIN can be used. The DC8 procedure shows potential as a time-efficient procedure, although more presentations than used in the current implementation are needed to reach the same accuracy as the D23 procedure. Consequently, the test time will increase. Furthermore, the average DC8 SRT is higher than the D23 SRT; thus, a different cutoff SRT must be determined. Finally, the DSS procedure is not useful as a screening test because of its poor test characteristics. Studies in which the potentially efficient new DIN procedures are compared to pure-tone thresholds are needed to validate these procedures.

Acknowledgments

This study was funded with the assistance from the Andrew W. Mellon Foundation's grant entitled "Supra- Institutional Initiative on the Advancement of Black South Africans Within the Arts, Humanities, and Social Sciences" (Grant G-41500687).

Disclosure Statement

De Wet Swanepoel has equity interest and consultancy and potential royalties in the hearX Group. The other authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.

References

Bergman, M. (1971). Hearing and aging: Implications of recent research findings. *Audiology, 10*(3), 164–171. https://doi.org/ 10.3109/00206097109072554

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 antiphasic stimuli. *Ear and Hearing, 41*(2), 442–450. https://doi.org/10.1097/AUD.0000000000000775

Denys, S., Hofmann, M., van Wieringen, A., & Wouters, J. (2019). Improving the efficiency of the digit triplet test using digit scoring with variable adaptive step sizes. *International Journal of Audiology, 58*(10), 670–677. https://doi.org/10.1080/14992027.2019.1622042

Global System for Mobile Communications Association. (2019). The mobile economy: Sub-Saharan Africa 2019. https://www.gsma.com/subsaharanafrica/resources/the-mobile-economysub-saharan-africa-2019

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. *International Journal of Audiology, 49*(5), 378–387. https://doi.org/10.3109/14992020903431272

Koole, A., Nagtegaal, A. P., Homans, N. C., Hofman, A., Baatenburg de Jong, R. J., & Goedegebure, A. (2016). Using the digits-in-noise test to estimate age-related hearing loss. *Ear and Hearing, 37*(5), 508–513. https://doi.org/10.1097/aud.0000000000000282

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., . . . Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet, 396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/ BM.2012.031

Moore, A. N., Rothpletz, A. M., & Preminger, J. E. (2015). The effect of chronological age on the acceptance of Internet-based hearing health care. *American Journal of Audiology, 24*(3), 280–283. https://doi.org/10.1044/2015_AJA-14-0082

Moore, D. R., Edmondson-Jones, M., Dawes, P., Fortnum, H., McCormack, A., Pierzycki, R. H., & Munro, K. J. (2014). Relation between speech-in-noise threshold, hearing loss and cognition from 40–69 years of age. *PLOS ONE, 9*(9), Article e107720. https://doi.org/10.1371/journal.pone.0107720

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. *International Journal of Audiology, 55*(7), 405–411. https://doi.org/10.3109/14992027.2016.1172269

Potgieter, J. M., Swanepoel, D. W., & Smits, C. (2018). Evaluating a smartphone digits-in-noise test as part of the audio- metric test battery. *South African Journal of Communication Disorders, 65*(1), 1–6. https://doi.org/10.4102/sajcd.v65i1.574

Saunders, G. H., & Cienkowski, K. M. (2002). A test to measure subjective and objective speech intelligibility. *Journal of the American Academy of Audiology, 13*(1), 38–49. https://doi.org/ 10.1055/s-0040-1715946

Saunders, G. H., Forsline, A., & Fausti, S. A. (2004). The Performance-Perceptual Test and its relationship to unaided reported handicap. *Ear and Hearing, 25*(2), 117–126. https://doi.org/10.1097/01.AUD.0000120360.05510.E5

Smits, C. (2017). Improving the efficiency of speech-in-noise hearing screening tests. *Ear and Hearing, 38*(6), e385–e388. https://doi.org/10.1097/AUD.0000000000000446

Smits, C., Goverts, S. T., & Festen, J. M. (2013). The digits-in- noise test: Assessing auditory speech recognition abilities in noise. *The Journal of the Acoustical Society of America, 133*(3), 1693–1706. https://doi.org/10.1121/1.4789933

Smits, C., & Houtgast, T. (2006). Measurements and calculations on the simple up-down adaptive procedure for speech-in-noise tests. *The Journal of the Acoustical Society of America, 120*(3), 1608–1621. https://doi.org/10.1121/1.2221405

Smits, C., Kapteyn, T. S., & Houtgast, T. (2004). Development and validation of an automatic speech-in-noise screening test by telephone. *International Journal of Audiology, 43*(1), 15–28. https://doi.org/10.1080/14992020400050004

Smits, C., Merkus, P., & Houtgast, T. (2006). How we do it: The Dutch functional hearing–screening tests by telephone and Internet. *Clinical Otolaryngology, 31*(5), 436–440.

https://doi.org/10.1111/j.1749-4486.2006.01195.x

Smits, C., Watson, C. S., Kidd, G. R., Moore, D. R., & Goverts, S. T. (2016). A comparison between the Dutch and American-English digits-in-noise (DIN) tests in normal-hearing listeners. *International Journal of Audiology, 55*(6), 358–365. https://doi.org/10.3109/14992027.2015.1137362

Swanepoel, D. W. (2017). Smartphone-based national hearing test launched in South Africa. *The Hearing Journal, 70*(1), 14–16. https://doi.org/10.1097/01.HJ.0000511726.41335.83

Swanepoel, D. W., De Sousa, K. C., Smits, C., & Moore, D. R. (2019). Mobile applications to detect hearing impairment: Opportunities and challenges. *Bulletin of the World Health Organization, 97*(10), 717–718. https://doi.org/10.2471/BLT.18.227728

Swanepoel, K. C. (2018). *Hearing screening with the digits-in-noise test: Evaluating a national test and new stimulus approach* [Master's thesis, University of Pretoria]. UPSpace Institutional Repository. https://repository.up.ac.za/handle/2263/70505

Van den Borre, E., Denys, S., van Wieringen, A., & Wouters, J. (2021). The digit triplet test: A scoping review. International *Journal of Audiology. Advance online publication.* https://doi.org/10.1080/14992027.2021.1902579

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology, 23*(10), 779–788. https://doi.org/10.3766/jaaa.23.10.4

Watson, C. S., Kidd, G. R., Miller, J. D., Smits, C., & Humes, L. E. (2012). Telephone screening tests for functionally im- paired hearing: Current use in seven countries and development of a U.S. version. *Journal of the American Academy of Audiology, 23*(10), 757–767. https://doi.org/10.3766/jaaa.23.10.2

Willberg, T., Buschermöhle, M., Sivonen, V., Aarnisalo, A. A., Löppönen, H., Kollmeier, B., & Dietz, A. (2016). The develop- ment and evaluation of the Finnish digit triplet test. *Acta Oto-Laryngologica, 136*(10), 1035–1040. https://doi.org/10.1080/00016489.2016.1175662

World Health Organization. (2018a). *Deafness prevention: Estimates*. https://www.who.int/deafness/estimates/en/

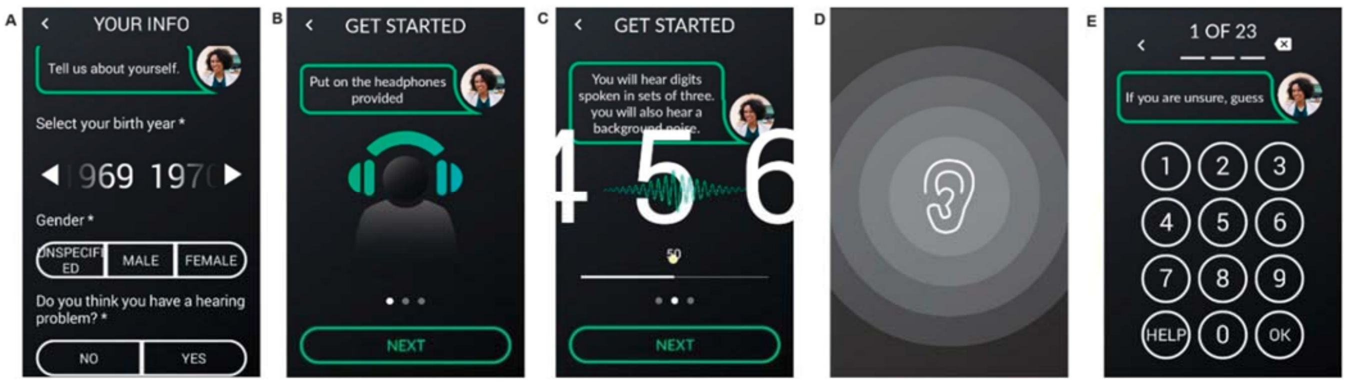
World Health Organization. (2018b). *WHO launches the hear- WHO app for mobile devices to help detect hearing loss*. www.who.int/news/item/01-03-2018-who-launches-the-hearwho-app-for-mobile-devices-to-help-detect-hearing-loss

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. *International Journal of Audiology, 51*(9), 697–707. https://doi.org/10.3109/14992027.2012.690078

Appendix (p. 1 of 4)

Screenshots From Smartphone Digits-in-Noise Applications

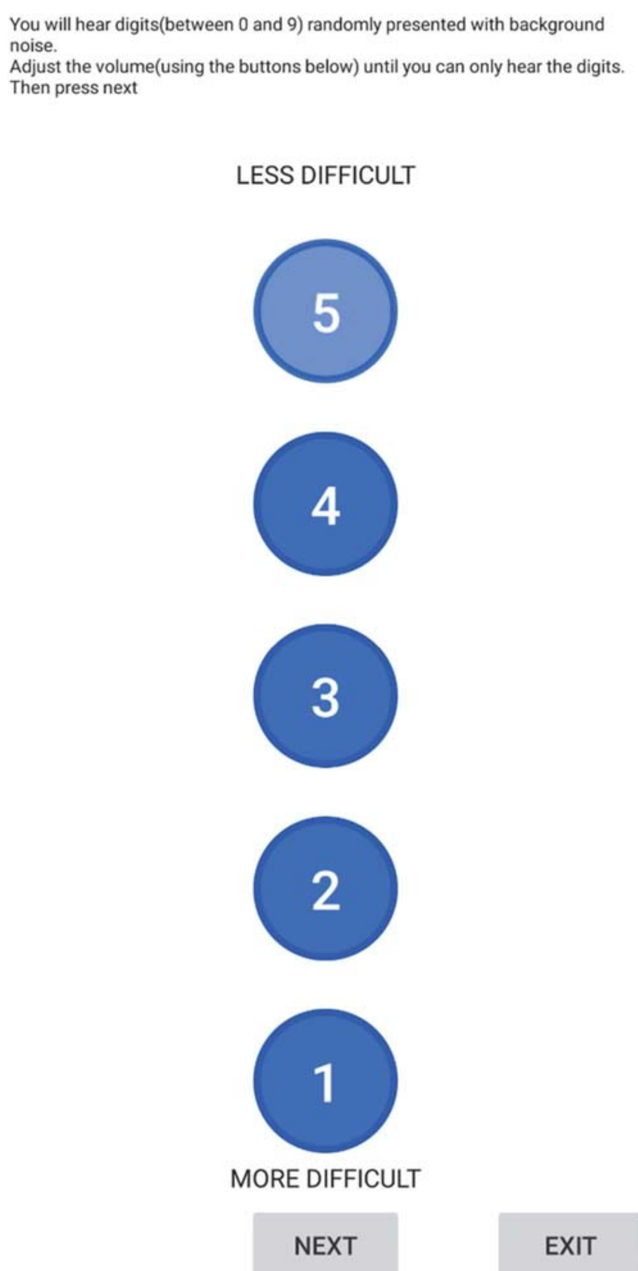
Figure A1. On-screen D23 test setup from the application. (A) Demographic information of participant. (B) Instructions to participant. (C) Setting of comfortable listening intensity. (D) Three digits presented in background noise. (E) Keypad to enter the three digits heard.



Appendix (p. 2 of 4)

Screenshots From Smartphone Digits-in-Noise Applications

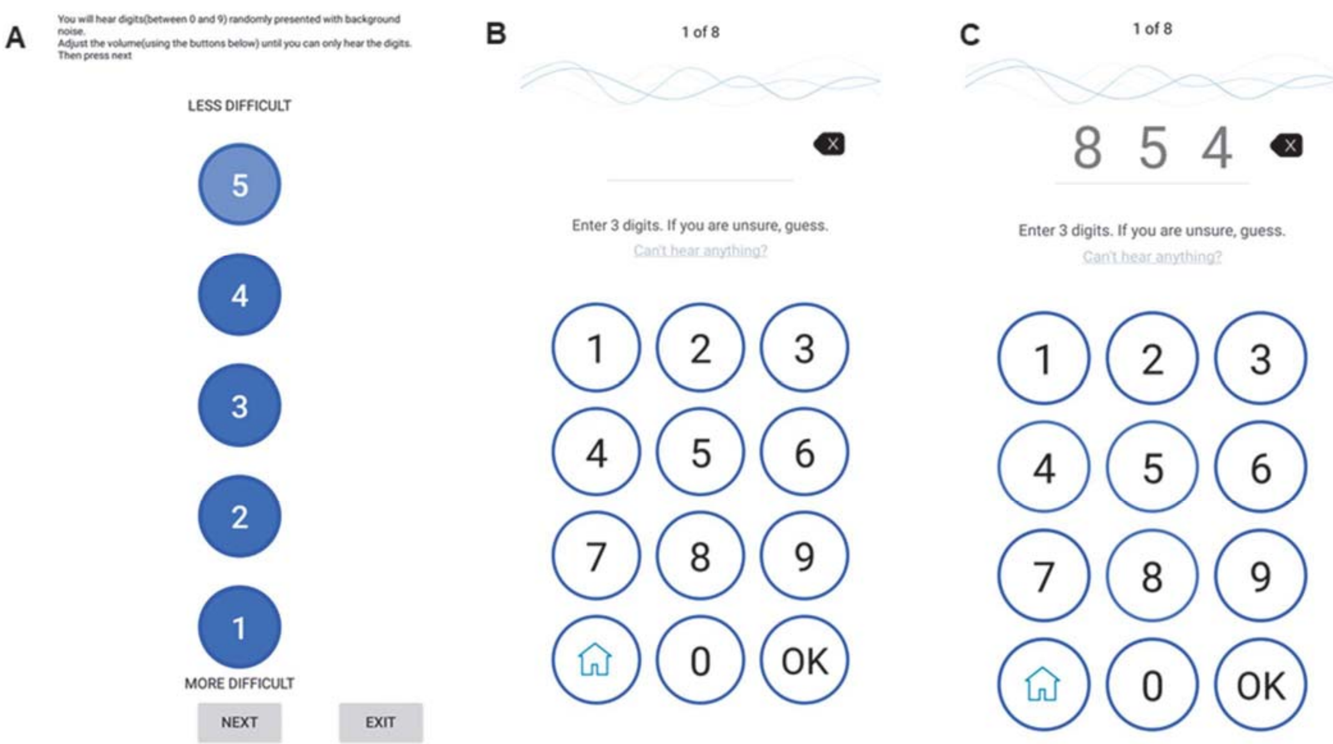
Figure A2. On-screen setup of the DSS slider.



Appendix (p. 3 of 4)

Figure A3. On-screen DC8 test setup from the application.

Phase 1: (A) Participant is instructed to adjust the volume using the buttons labeled 1–5 until they can only hear the digits and then press "Next."
Phase 2: (B) Three digits are presented in background noise.
(C) Participants type in the three digits heard.



Appendix (p. 4 of 4)

Figure A4. On-screen DF test setup from the application.

(A) Setting of comfortable listening intensity (right to increase, left to decrease).
(B) Three digits are presented in background noise.
(C) Participant types in the three digits heard on the keypad provided.