Comparison of Speech Reception Thresholds for diotic, dichotic and antiphasic headphone presentations of digits-in-noise triplets using Dantale I material

Palle Rye[1,*], Rasmus Overgaard Andersen[1], Gert Ravn[2], Dorte Hammershøi[1]

[1]Department of Electronic Systems, AI and Sound, Aalborg, Denmark.

[2] Teknisk Audiologisk Laboratorium (TAL), FORCE Technology, Odense, Denmark.

*par@es.aau.dk

Abstract

A complete and accurate diagnosis is the prerequisite of efficient treatment of any disorder. For hearing disorders, the primary diagnostic tools for characterizing the type and severity typically include a pure-tone audiogram describing the sensitivity loss **across** the frequency range typical for speech and one or more speech recognition tests. While there is typically a high level of correlation between the audiogram and the speech recognition tests, the speech test may reveal different characteristics of the hearing disorders. The research interest in out-of-clinic versions of speech in noise has increased in recent years. Different implementations exist in various languages, e.g. the digits-in-noise triplet test made available by the World Health Organization. The present study examines a similar triplet test featuring digits in Danish originating from the DANTALE I material. The original masking noise of the DANTALE I material did not contain sufficient high-frequency content to mask the digits effectively and equally well, so a new masking noise was derived from the digit material. Three different spatial configurations of the presented speech in noise were tested: diotic speech in dichotic noise, and antiphasic speech in both diotic and dichotic noise. Initial test on 18 subjects (5 subjects with better hearing ear pure-tone average (PTA_{BE}) > 30dB) in the age range 53-81 years (mean 64.4 years) shown reasonable correlation (r > 0.86 for all three configurations) between PTA_{BE} and the estimated speech reception thresholds.

Keywords: speech audiometry, digits-in-noise, antiphasic speech

1 Introduction

For decades air conducted pure tone audiometry has remained the gold standard in screening and the mainstay in diagnosing hearing disorders. Methods, equipment and environment requirements are standardised[1], and are widely accepted to provide useful, accurate and repeatable results. Typically performed in a clinical environment, it relies on a well-controlled acoustic environment with low noise levels, regularly calibrated transducers, and hearing health-care professionals for administering the test with suitable patient guidance. Although many implementations of automatic procedures for estimating the pure tone threshold exist, some even self-administered, it has yet to gain widespread in-clinic use. The goal of hearing rehabilitation is predominately improving speech understanding. A high level of correlation between the pure tone thresholds and the performance in speech recognition tests allows either to be used as an estimator for the other. Many factors influence speech perception[2], from the type of speech material, the presenting speaker, acoustic environment, hearing characteristics and cognitive abilities of the subject, response collection and scoring method. As with pure tone threshold audiometry, the basic methods for speech recognition tests are standardised[3], with an important difference of being inherently language-dependent.

In current Danish clinical practice, two speech audiometric procedures are routinely used[4]: the supra-threshold Word Recognition Score (WRS), counting the percentage of words discerned correctly and the Speech Reception Threshold (SRT), estimating the speech level where the test subject understands at least half of the presented material. The speech material for these test use lists of unrelated monosyllabic words from DANTALE I [5] and are typically performed in quiet without competing noise. The change in performance as a function of intensity change is described by the intelligibility slope, regarded as a smooth function from zero to full intelligibility over a range of intensities. The intelligibility slope steepness has been found to increase when a competing noise is presented simultaneously with speech sentences[6], [7]. Annex C of ISO8253-3[3] lists typical reference SRT using example data from [7], where the steepness of the intelligibility slope at the 50% point was 11%/dB in quiet and 19%/dB in competing noise. Sentence lists are more complex than monosyllabic word lists, and it is challenging to compile sentence lists that provide equivalent results. Notable examples of this type of hearing-in-noise-tests (HINT) for different languages include English[8], Swedish[9], Danish[10], and German[11]. Currently, HINT relies on operator interaction for response collection and scoring, inhibiting its use in a self-administered scenario.

The digits-in-noise (DIN) test, or digit-triplet test, was developed as a self-administered speech-in-noise test [12], which sought to screen the Dutch population over landline phones[13]. Essentially a measure of the SRT in competing noise, the suprathreshold test does not rely on carefully calibrated transducers or low ambient noise levels yet exhibits high test-retest reliability and high correlation with pure tone threshold average (PTA) calculated from four frequencies 0.5, 1k, 2k and 4kHz. Following the initial Dutch advances, the DIN test was adapted to different languages, targeted populations, and ubiquitous test platforms, such as smartphones apps and internet browsers. A recent scoping review[14] provides a comprehensive overview and notes the intelligibility slope steepness being in the range of 15-20%/dB across most of the reviewed studies and language adaptations. The earliest DIN studies tend to use the monaural presentation of both speech and noise sequentially testing each ear. Later studies used diotic tests with the same stimuli simultaneously in both ears (the nomenclature N_0S_0 signifies that there are no interaural differences for neither the noise, N, nor the stimuli, S, in this case, speech). While the diotic test halves the test time compared to each ear in succession, it relies strongly on the better hearing ear, and asymmetric hearing losses may be missed in the screening. A dichotic noise, N_u, timewise completely uncorrelated between the two ears yet has the same long-term spectrum as diotic speech S_0, appears to have different spatial characteristics than a diotic signal fully correlated between the ears. For the same volume setting, binaural summation of loudness means N_u is perceived to be louder than N_0. Nevertheless, due to binaural unmasking, the reception threshold of a N_uS_0 spatial configuration can be expected to be a few dB lower than N_0S_0. Such threshold improvement due to binaural unmasking is called the binaural masking level difference (BMLD)[15]. For diotic noise combined with antiphasic stimuli, N_0S_π, meaning the stimuli is in opposite phase between ears, binaural unmasking also enhances the stimuli detection relative to $N_0 S_0$ resulting in BMLD. Recent studies have found a higher correlation between worse hearing ear PTA (PTA_{WE}) and antiphasic DIN, which improves the screening sensitivity to asymmetric hearing impairment[16]. A mobile application featuring antiphasic DIN was adopted by the World Health Organization[17].

In recent years, the interest in eHealth research and out-of-clinic hearing healthcare has been surging[18]. Examination of the performance of supervised clinical methods versus self-administered methods in ecologically valid out-of-clinic settings can help reveal potential critical issues and benefits of out-of-clinic approaches. Previous research established several benefits of favouring antiphasic speech over the diotic speech with diotic noise. The primary goal of the present study is to examine differences between three other spatial configurations to determine any potential advantages of using dichotic uncorrelated noise. Secondly, it is of interest whether some spatial configurations of DIN correlate with HINT measurements encouraging further research into applying unsupervised DIN speech audiometry as a diagnostic measure or even as hearing aid fitting validation tool, duly noting that relying on presenting antiphasic speech could prove challenging in the latter case.

2 Methods and Materials

2.1 Participants

Eighteen listeners (8 female) aged between 53 and 81 years (mean 64.4 years) participated in the study. Four participants were experienced hearing aid users, two with moderate impairment (PTA 35-50 dB)[19], one with moderate-to-severe (PTA 50-65dB) and one with strongly asymmetric hearing impairment. Eleven participants had normal hearing (PTA<20), one with moderate hearing impairment and the remaining with mild hearing impairment (PTA 20-35dB).

2.2 Test platform and setup

The DIN test was programmed in C#, compiled on the Unity platform and installed on an Asus Zenpad 3S10 Android tablet. The participants entered age and gender data and answered a 12-item computer self-efficacy questionnaire before starting with the DIN tests, followed by various other self-administered audiological tests and questionnaires. A familiarisation phase with five triplet presentations preceded data collection of the first spatial configuration. The total test time of the complete test battery including exit interviews was designed to be conducted in less than one hour per participant. A set of active noise-cancelling headphones BOSE QC35II connected to the tablet via Bluetooth was used for presenting the stimuli. All tests were conducted in an out-of-clinic environment typical to the intended use case, and ambient noise levels were monitored. A rest period was provided between the self-administered test battery and the more traditional audiometric tests. For eight participants the tests were conducted on the same day after the rest period, while the other ten had a minimum of two weeks between the self-administered and traditional tests.

   Pure tone thresholds were measured manually according to ISO 8253-1[1] using the audiometer functionality of the InterAcoustics Affinity Test Suite connected to a RadioEar DD45 headphone. The HINT was evaluated using a set of MATLAB scripts from the original author of the Danish procedure[10]. The scripts were executed on a laptop controlling a soundcard RME Fireface UCX connected to Sennheiser HDA200 headphones. Headphones for all experiments were calibrated using an artificial ear coupler B&K Type 4153.

2.3 Speech material and adaptive procedure

Seven monosyllabic numbers (0, 1, 2, 3, 5, 6 and 7) were extracted from DANTALE I list and recombined into random digit triplet sequences. Each triplet sequence contains three different digits, and the time between each digit is 0.5 seconds. The playback noise level was fixed to 60dB SPL throughout the test unless the adaptive algorithm should converge towards providing a positive SNR above +20dB. In such a case the noise level would be reduced instead to avoid unnaturally loud speech levels. The initial SNR was set to 0dB. An adaptive 1up-1down procedure was used for each correct triplet. A digit-scoring approach was used where a triplet was considered correct when two out of the three digits were correctly identified. Antiphasic speech was simple phase-reversal of one channel. Due to BMLD, considerable variation in SRT across the three spatial conditions was likely to occur. Indeed pilot tests revealed SRT differences in excess of 10dB between the three spatial conditions. A procedure with an adaptive step size was chosen to start from the same initial SNR across the spatial conditions, where the initial step size of the SNR was set to 5dB to allow fast descent toward the threshold. A total of 8 reversals was chosen as a compromise between a manageable number of trials for each spatial condition and an adequate number of trials around the threshold. Following each of the first three reversals the step size was reduced by 1dB, with three reversals of 2dB step size before finishing 1dB for the final two reversals. The threshold was calculated as the average of the last five reversals.

2.4 Masking noise

Following analysis of psychometric curves of individual digits in a pilot test, changing the speech-shaped masking noise from DANTALE I was deemed necessary, since digits 3, 6 and 7 were recognisable at

significantly lower SNRs than the other digits. A similar issue was reported earlier for the DANTALE I sentence material[20], recommending that certain word lists be avoided when using DANTALE I masking noise. For the present study a new speech-shaped noise was generated by filtering Gaussian random noise to have the same long-term spectrum as a sequence of all digits, thus containing more high-frequency energy. Uncorrelated channels were achieved by using random playback points in the noise wave file separated by at least 5 seconds. Third-octave analysis of the speech and noise signals are shown in Figure 1 for DANTALE I and Figure 2 for the noise used in this study.
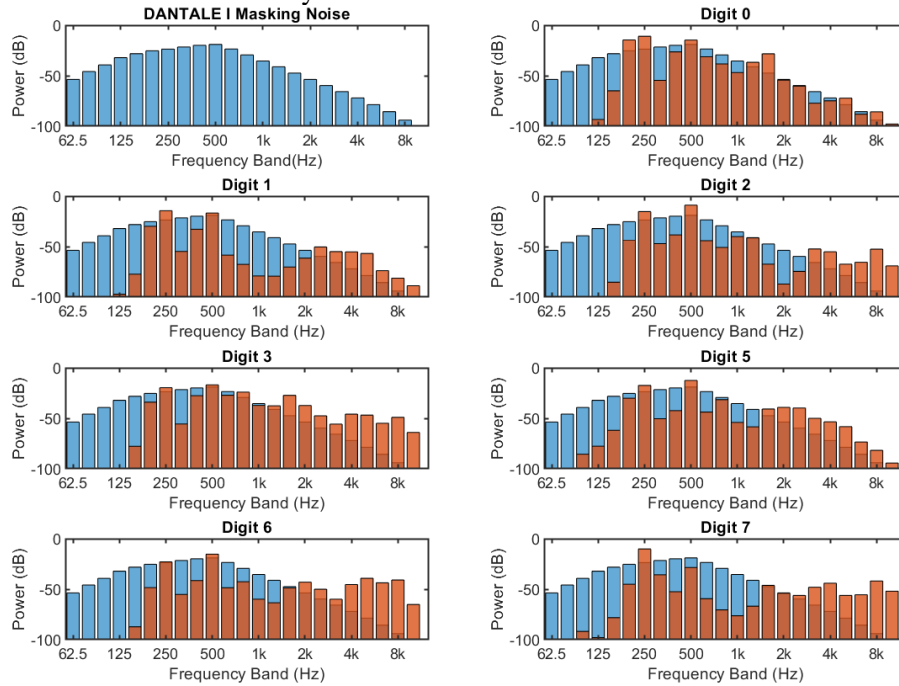

Figure 1: Third-octave spectra of the individual digits with the masking noise from DANTALE I
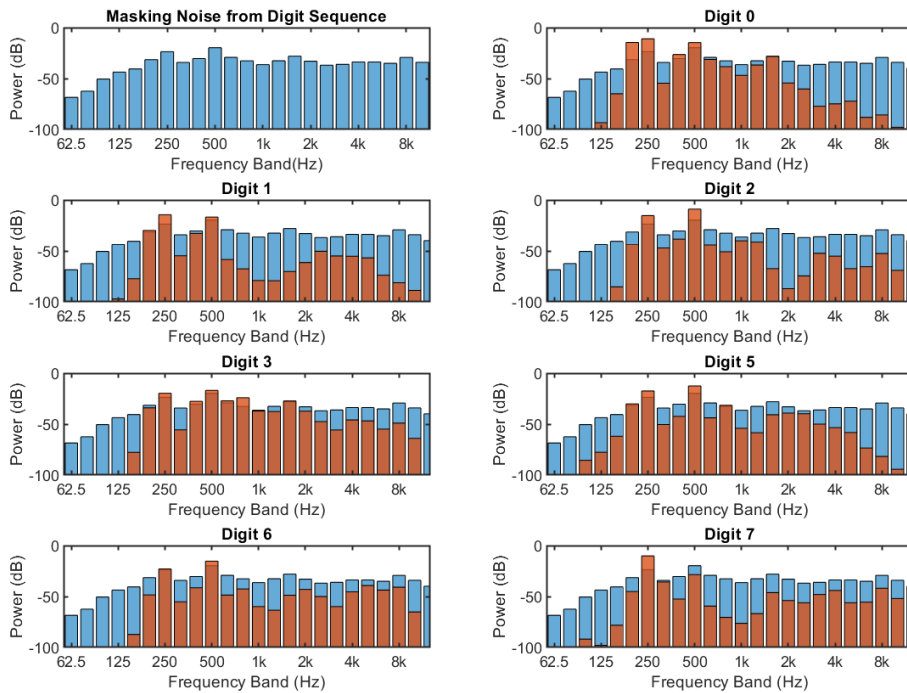

Figure 2: Third-octave spectra of the individual digits with the speech shaped noise used in this study

3   Results

The average number of digit triplet trials for determining the SRT was 15.3 (SD=2.9). The total trial time for the three spatial conditions was, on average 6.6 (SD=1.4) minutes. Two typical examples of the adaptive staircase for the three spatial conditions are shown in Figure 3.
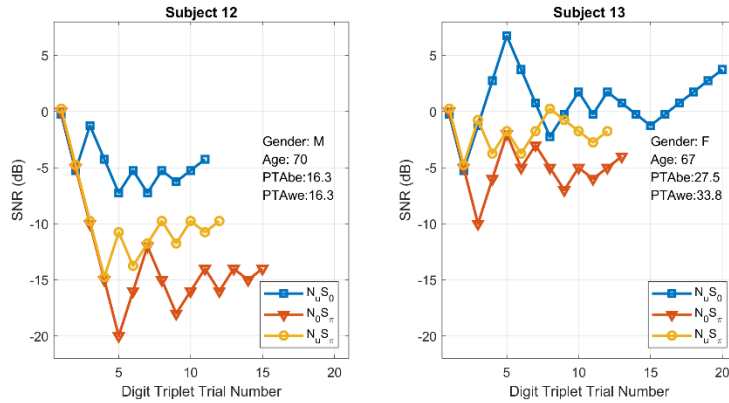

Figure 3: Typical examples of the adaptive staircases for the three spatial configurations.

Linear regression fit including 95% confidence intervals (CI) are shown in Figure 4, where each column represents a spatial condition so that BMLD may be revealed as a systematic offset across columns. Intercepts of N_uS_0 and N_uS_π differ by 2.3 dB for PTA_{BE} and 2.8 dB for PTA_{WE}. The middle column N_0S_π is clearly offset downwards compared to the other conditions and features a steeper regression slope allowing for greater separation of PTA in a screening scenario. Top row (better hearing ear) regression features steeper slopes than bottom row (worse hearing ear) and higher correlation in two of three conditions. Regressed DIN SRT from PTA (black lines of Figure 4) illustrates potential separation of normal hearing (<20dB), from mild (20-35 dB), moderate (35-50dB) and moderate-severe (50-65dB) hearing configurations.
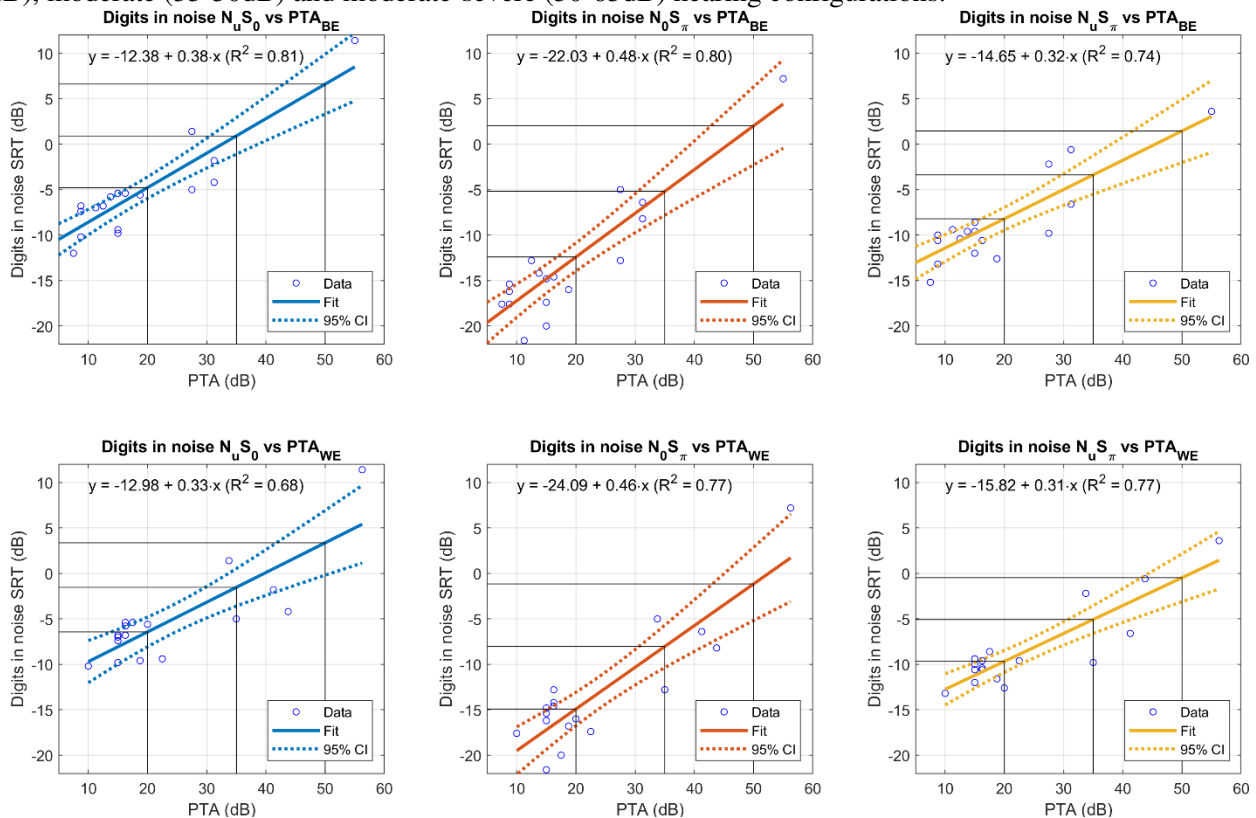

Figure 4: Linear regression plots with PTA as predictor and DIN SRT as the response variable.

Linear regression relating DIN SRT for all three spatial conditions across columns to HINT thresholds for diotic $HINT_{Diotic}$ (N_0S_0), better hearing ear ($HINT_{BE}$) and worse hearing ear ($HINT_{WE}$) across rows is shown in Figure 5. Shallow slopes are generally observed along with relatively poor correlation over narrow ranges (5.8, 5.9 and 7.4 dB) of HINT SRT with sizable variance (1.9, 3.4 and 5.0 dB) determined for this population. The steepest slopes and highest correlations are found in the bottom row featuring the $HINT_{WE}$. Here the groups with and without hearing impairment also appear more clearly separated.
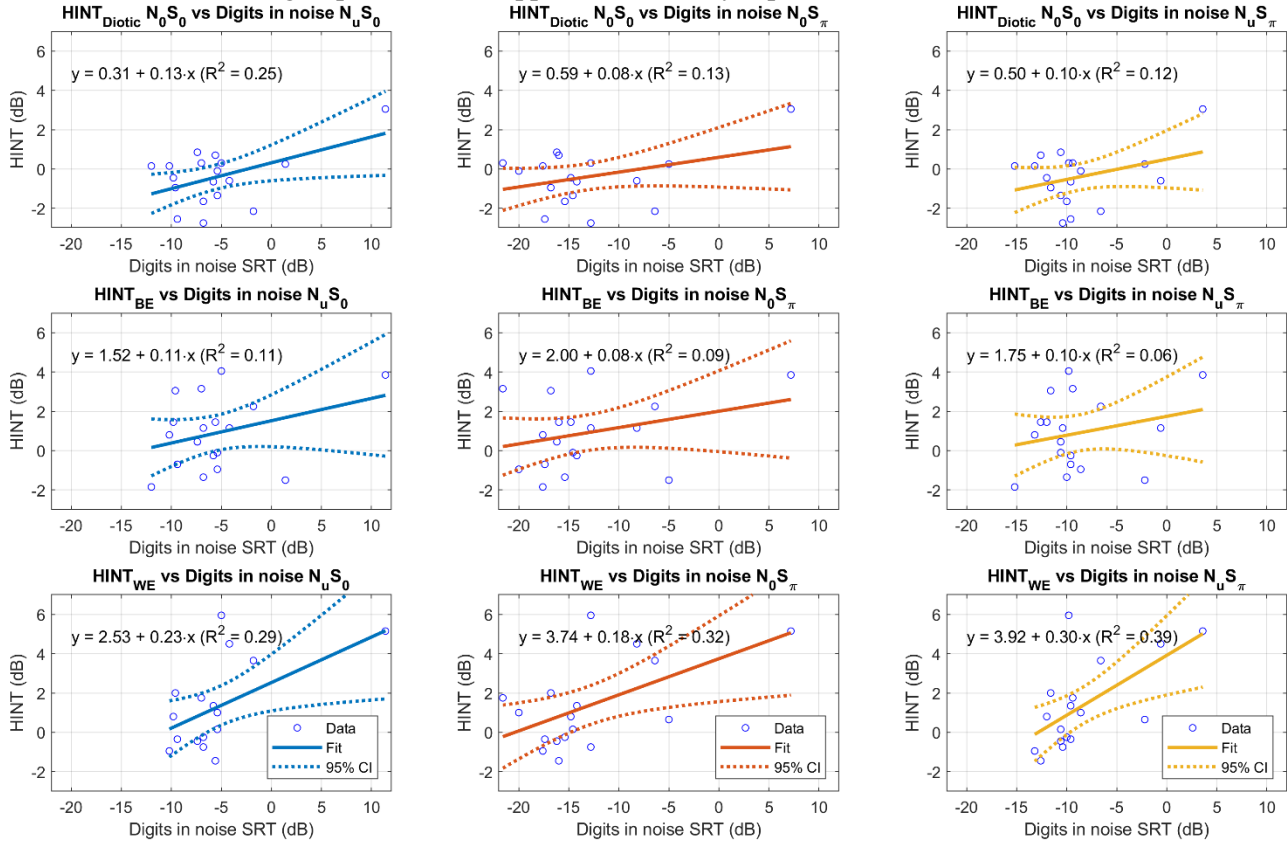

Figure 5: Linear regression plots with DIN SRT as predictor and HINT as the response variable

The correlation matrix is shown in Table 1. The studied population has significant PTA symmetry between ears ($r = 0.93$) and between each of the three spatial conditions ($r < 0.85$). Significant correlations are found between all DIN spatial configurations and PTA_{BE} ($r > 0.86$) and PTA_{WE} ($r > 0.83$). The $HINT_{WE}$ correlates significantly with PTA_{WE} ($r = 0.79$), while the $HINT_{BE}$ does not correlate significantly with PTA_{BE} ($r = 0.35$).

Table 1: Correlation between PTA at the better and worse hearing ear, the measured DIN SRT levels for the three spatial configurations, and HINT thresholds for better hearing ear, worse hearing ear and diotic condition; Correlations in bold are significant at the 0.001 level.

	PTA_{BE}	PTA_{WE}	N_uS_0	N_0S_π	N_uS_π	$HINT_{Diotic}$	$HINT_{BE}$	$HINT_{WE}$
PTA_{BE}	1							
PTA_{WE}	**0.93**	1						
N_uS_0	**0.90**	**0.83**	1					
N_0S_π	**0.89**	**0.88**	**0.90**	1				
N_uS_π	**0.86**	**0.88**	**0.87**	**0.85**	1			
$HINT_{Diotic}$	0.47	0.35	0.56	0.39	0.41	1		
$HINT_{BE}$	0.35	0.41	0.24	0.27	0.14	0.38	1	
$HINT_{WE}$	0.68	**0.79**	0.54	0.56	0.62	0.32	0.64	1

4 Discussion

For the sake of reducing test time in a large test battery, short trial lengths (average of 15.3 triplet trials per spatial configuration) were achieved by using an adaptive procedure that was based on 8 reversals and reducing step size. The correlation between SRT and PTA was reported to change only slightly after 15 trials of 2dB in an earlier study[21]. A recent study[22] reported SRT test-retest differences for adults of 0.98 dB (SD = 2.91) with only 6 reversals and an average of 23.2 (SD = 4.1) trials with an adaptive step size changing from 6 dB initially to 3 dB after first reversal. The favourable number of trials per reversal was achieved by changing from a 1-down-1-up to a 2-down-1-up paradigm after the first two reversals.

A similar approach could be adopted for future implementations of the DIN in the present study. This could prove beneficial, especially if the slopes of the psychometric function for the DIN are found to be less steep for individuals of a larger population than examined here. Various impairments and characteristics of hearing may be imagined where the SRT estimation of a test subject would benefit from a more trials around threshold. This could also be achieved by selecting more than 8 reversals, fixing a minimum number of trials, or using a stopping criteria based on observed individual threshold variance[23]. The high correlation between antiphasic speech in diotic noise and PTA_{WE} (r = 0.88) is similar to the findings of antiphasic DIN studies in other languages, e.g. South-African English (r = 0.82)[16] and French (r = 0.82)[24].

The observed offset between intercepts of N_uS_0 and N_uS_π spatial configurations of 2.3 dB and 2.8 dB for better and worse hearing ear respectively are in agreement with a BMLD of about 2 dB described in [15] for tonal stimuli in uncorrelated noise. However, uncorrelated noise, N_u, yielded flatter regression slopes with PTA in both conditions compared to N_0 suggesting no benefit for screening purposes. Due to the steeper slopes found in the correlation with PTA, the present study suggest antiphasic speech in diotic noise (N_0S_π) is more appropriate for screening than the other spatial configurations.

For diagnostic purposes it may be beneficial to find a measure that is not directly correlated with PTA yet may reveal other deficits than loudness recruitment. A sentence-based HINT, arguably a more ecologically valid candidate, might suit such purpose for in-clinic use. With DIN as a potential candidate for self-administered out-of-clinic diagnostic test any significant correlation with HINT would encourage further research into the suitability. However, low correlation between DIN and HINT measures was observed for the tested population, with possible exception for DIN and HINT worse hearing ear performance. The HINT SRT for the tested population exhibit sizable variance up to over a relatively narrow range (up to 5.0 dB over a 7.4 dB range). This could be indicative of very steep psychometric slopes for sentence-based paradigm, where a complete trial fails even when getting single word wrong (except for tense, order and so on). In HINT some test subjects might not feel entirely comfortable with guessing and opt to pass if they do not understand the complete sentence. Even when one or more digits are not understood, the DIN method forces the subject to guess the unheard digits in order to proceed. Inspection of responses well below threshold revealed how some subjects entered triplets, which could be interpreted as opting to pass a specific trial: e.g. 1-2-3 or 0-0-0 in case of completely missing a triplet.

5 Conclusions

The goal of the present study was to examine differences between three spatial configurations of digits in noise and determine potential advantages of using dichotic uncorrelated noise with diotic or antiphasic speech. For screening purposes no advantages of dichotic uncorrelated noise were found. In line with previous research, antiphasic speech in diotic noise exhibited significant correlation worse ear PTA and remains the strongest candidate of the three examined spatial configurations for screening purposes. The study did not find convincing correlation between monosyllabic DIN and sentence-based HINT measurements, however the research interest in using unsupervised DIN speech audiometry as a diagnostic measure or hearing aid fitting validation tool remains.

Acknowledgements

References

[1] ISO 8253-1, *Acoustics – Audiometric test methods – Part 1 : Pure-tone air and bone conduction audiometry*. International Organization for Standardization, 2010.

[2] Working Group on Speech Understanding, "Speech understanding and aging," *J. Acoust. Soc. Am.*, vol. 83, no. 3, pp. 859–895, Mar. 1988, doi: 10.1121/1.395965.

[3] ISO 8253-3, *Acoustics – Audiometric test methods – Part 3: Speech audiometry*. International Organization for Standardization, 2022.

[4] A. Wolff, *Health-Related Quality of Life Following Hearing Aid Treatment a large Cohort study*. Aalborg Universitetsforlag, 2019.

[5] C. Elberling, C. Ludvigsen, and P. E. Lyregaard, "Dantale: A New Danish Speech Material," *Scand. Audiol.*, vol. 18, no. 3, pp. 169–175, Jan. 1989, doi: 10.3109/01050398909070742.

[6] R. Plomp and A. M. Mimpen, "Improving the Reliability of Testing the Speech Reception Threshold for Sentences," *Int. J. Audiol.*, vol. 18, no. 1, pp. 43–52, Jan. 1979, doi: 10.3109/00206097909072618.

[7] B. Kollmeier and M. Wesselkamp, "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," *J. Acoust. Soc. Am.*, vol. 102, no. 4, pp. 2412–2421, Oct. 1997, doi: 10.1121/1.419624.

[8] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994, doi: 10.1121/1.408469.

[9] M. Hällgren, B. Larsby, and S. Arlinger, "A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition," *Int. J. Audiol.*, vol. 45, no. 4, pp. 227–237, Jan. 2006, doi: 10.1080/14992020500429583.

[10] J. B. Nielsen and T. Dau, "The Danish hearing in noise test," *Int. J. Audiol.*, vol. 50, no. 3, pp. 202–208, Mar. 2011, doi: 10.3109/14992027.2010.524254.

[11] J. Joiko, A. Bohnert, S. Strieth, S. D. Soli, and T. Rader, "The German hearing in noise test," *Int. J. Audiol.*, vol. 60, no. 11, pp. 927–933, Nov. 2021, doi: 10.1080/14992027.2020.1837969.

[12] C. Smits, T. S. Kapteyn, and T. Houtgast, "Development and validation of an automatic speech-in-noise screening test by telephone," *Int. J. Audiol.*, vol. 43, no. 1, pp. 15–28, Jan. 2004, doi: 10.1080/14992020400050004.

[13] C. Smits, P. Merkus, and T. Houtgast, "How we do it : The Dutch functional hearing – screening tests by telephone and internet," *Clin. Otolaryngol.*, vol. 31, no. 5, pp. 436–440, 2006, doi: https://doi.org/10.1111/j.1749-4486.2006.01195.x.

[14] E. Van den Borre, S. Denys, A. van Wieringen, and J. Wouters, "The digit triplet test: a scoping review," *Int. J. Audiol.*, vol. 60, no. 12, pp. 946–963, Dec. 2021, doi: 10.1080/14992027.2021.1902579.

[15] J. F. Culling and M. Lavandier, "Binaural Unmasking and Spatial Release from Masking," in *Binaural Hearing*, vol. 73, R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper, Eds. Cham: Springer International Publishing, 2021, pp. 209–241. doi: 10.1007/978-3-030-57100-9_8.

[16] K. C. De Sousa, D. W. Swanepoel, D. R. Moore, H. C. Myburgh, and C. Smits, "Improving Sensitivity of the Digits-In-Noise Test Using Antiphasic Stimuli," *Ear Hear.*, vol. 41, no. 2, pp. 442–450, Mar. 2020, doi: 10.1097/AUD.0000000000000775.

[17] D. W. Swanepoel, K. C. De Sousa, C. Smits, and D. R. Moore, "Mobile applications to detect hearing impairment: opportunities and challenges," *Bull. World Health Organ.*, vol. 97, no. 10, pp. 717–718, Oct. 2019, doi: 10.2471/BLT.18.227728.

[18] A. Paglialonga, A. Cleveland Nielsen, E. Ingo, C. Barr, and A. Laplante-Lévesque, "eHealth and the hearing aid adult patient journey: A state-of-the-art review," *Biomed. Eng. Online*, vol. 17, no. 1, pp. 1–26, 2018, doi: 10.1186/s12938-018-0531-3.

[19] G. Stevens, S. Flaxman, E. Brunskill, M. Mascarenhas, C. D. Mathers, and M. Finucane, "Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries," *Eur. J. Public Health*, vol. 23, no. 1, pp. 146–152, Feb. 2013, doi: 10.1093/eurpub/ckr176.

[20] G. Keidser, "Normative Data in Quiet and in Noise for 'Dantale'—A Danish Speech Material," *Scand. Audiol.*, vol. 22, no. 4, pp. 231–236, Jan. 1993, doi: 10.3109/01050399309047474.

[21] C. S. Watson, G. R. Kidd, J. D. Miller, C. Smits, and L. E. Humes, "Telephone Screening Tests for Functionally Impaired Hearing: Current Use in Seven Countries and Development of a US Version," *J. Am. Acad. Audiol.*, vol. 23, no. 10, pp. 757–767, Nov. 2012, doi: 10.3766/jaaa.23.10.2.

[22] D. R. Moore *et al.*, "FreeHear: A New Sound-Field Speech-in-Babble Hearing Assessment Tool," *Trends Hear.*, vol. 23, p. 233121651987237, Jan. 2019, doi: 10.1177/2331216519872378.

[23] H. Dillon, E. F. Beach, J. Seymour, L. Carter, and M. Golding, "Development of Telscreen: a telephone-based speech-in-noise hearing screening test with a novel masking noise and scoring procedure," *Int. J. Audiol.*, vol. 55, no. 8, pp. 463–471, 2016, doi: 10.3109/14992027.2016.1172268.

[24] J.-C. Ceccato *et al.*, "French Version of the Antiphasic Digits-in-Noise Test for Smartphone Hearing Screening," *Front. Public Health*, vol. 9, p. 725080, Oct. 2021, doi: 10.3389/fpubh.2021.725080.