

**Document Version**

Final published version

**Licence**

CC BY

**Citation (APA)**

Knipscheer, B., Stronks, H. C., Briaire, J. J., & Frijns, J. H. M. (2026). Speech intelligibility in noise with everyday sentences correlates better with matrix sentences than with digits for cochlear implant users and typical hearing listeners. *International Journal of Audiology*. <https://doi.org/10.1080/14992027.2025.2607537>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.  
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

**Sharing and reuse**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



## Speech intelligibility in noise with everyday sentences correlates better with matrix sentences than with digits for cochlear implant users and typical hearing listeners

Bram Knipscheer, Hendrik Christiaan Stronks, Jeroen Johannes Briaire & Johan Hubertus Maria Frijns

To cite this article: Bram Knipscheer, Hendrik Christiaan Stronks, Jeroen Johannes Briaire & Johan Hubertus Maria Frijns (23 Jan 2026): Speech intelligibility in noise with everyday sentences correlates better with matrix sentences than with digits for cochlear implant users and typical hearing listeners, International Journal of Audiology, DOI: [10.1080/14992027.2025.2607537](https://doi.org/10.1080/14992027.2025.2607537)

To link to this article: <https://doi.org/10.1080/14992027.2025.2607537>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of British Society of Audiology, International Society of Audiology, and Nordic Audiological Society.



Published online: 23 Jan 2026.



Submit your article to this journal [↗](#)



Article views: 331







View related articles [↗](#)



View Crossmark data [↗](#)

# Speech intelligibility in noise with everyday sentences correlates better with matrix sentences than with digits for cochlear implant users and typical hearing listeners

Bram Knipscheer<sup>a</sup> , Hendrik Christiaan Stronks<sup>a,b</sup> , Jeroen Johannes Briaire<sup>a</sup>  and Johan Hubertus Maria Frijns<sup>a,b,c</sup> 

<sup>a</sup>Department of Otorhinolaryngology and Head & Neck Surgery, Leiden University Medical Center, Leiden, the Netherlands; <sup>b</sup>Leiden Institute for Brain and Cognition, Leiden, the Netherlands; <sup>c</sup>Department of Bioelectronics, Delft University of Technology, Delft, the Netherlands

## ABSTRACT

**Objective:** To determine correlations between speech recognition thresholds (SRTs) for everyday sentences compared to Matrix sentences and digits-in-noise (DIN) triplets.

**Design:** Comparative analysis of SRTs across three speech materials in both cochlear implant (CI) users and typical hearing (TH) listeners, using linear regression and analysis of covariance.

**Study Sample:** 18 experienced CI users (mean age 63 ± 5 years) and 18 age-equivalent TH listeners (mean age 62 ± 12 years), all naive to the test materials.

**Results:** SRTs of Matrix sentences and everyday sentences correlated significantly ( $R^2$ : 0.81 for CI, 0.71 for TH), as did SRTs of DIN triplets and everyday sentences ( $R^2$ : 0.42 for CI, 0.28 for TH). Regression slopes did not differ significantly between CI and TH groups in either comparison. However, intercepts differed significantly between the CI (−2.65) and TH (−6.70) groups for the DIN triplets, but not for the Matrix sentences. Slopes for the DIN triplets deviated significantly from unity for both groups.

**Conclusions:** Dutch/Flemish Matrix sentence SRTs closely correlate with everyday sentence SRTs for both CI users and TH listeners, establishing it as a reliable alternative for repeated assessment. DIN triplet SRTs showed weaker correlations with everyday sentences and with significant intercept differences between groups.

## ARTICLE HISTORY

Received 3 July 2025

Revised 5 December 2025

Accepted 15 December 2025

## KEYWORDS



Cochlear implants;  
sensorineural hearing loss;  
Matrix test; Leuven  
intelligibility sentences test;  
digits-in-noise

## Introduction

Cochlear implants (CIs) represent the first-line treatment for severe-to-profound sensorineural hearing loss. For CI users and other hearing-impaired populations, speech intelligibility constitutes the primary outcome parameter for auditory rehabilitation. Accurate speech perception testing in clinical practice and scientific research is therefore critical for understanding and optimising communication outcomes.

To effectively capture speech perception levels in daily life, the speech material of a listening test ideally should reflect everyday communication. Ecological validity—how well experimental conditions approximate real-world scenarios—represents a fundamental consideration in speech perception assessment (Dingemans

and Goedegebure 2019). An ecologically valid speech perception paradigm incorporates sentences delivered at conventional speech rates with typical articulatory clarity. These sentences usually contain contextual cues, allowing listeners to use redundant information to predict missed words. This contextual dimension is particularly relevant for CI users, who rely far more than individuals with typical hearing (TH) on contextual cues for speech comprehension (Dingemans and Goedegebure 2019; Winn 2016). By more accurately reflecting the linguistic and cognitive demands of everyday verbal communication, sentence-based tests with an open-set response format show superior ecological validity over word-based or closed-set alternatives (Billings et al. 2024; Dingemans and Goedegebure 2019).

**CONTACT** Hendrik Christiaan Stronks  h.c.stronks@lumc.nl  Department of Otorhinolaryngology and Head & Neck Surgery, Leiden University Medical Center, Leiden, the Netherlands.

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of British Society of Audiology, International Society of Audiology, and Nordic Audiological Society.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

The initial adaptation period following CI implantation necessitates intensive training and frequent testing to monitor progress. In research settings, speech perception tests are repeatedly administered to evaluate multiple experimental conditions. Although their reputation for validity has made open-set sentence-based tests the preferred paradigm for assessing speech perception, they entail two limitations that need consideration. First, sentence-based tests contain a finite number of unique stimuli. Everyday sentences are associated with content learning effects, meaning that participants potentially memorise the material when exposed to the same sentences repeatedly. Consequently, they cannot be repeatedly administered to the same participant without compromising measurement validity, with this limitation persisting for at least six months (Wagener and Brand 2005; Yund and Woods 2010). By contrast, random word or digit combinations demonstrate negligible content learning effects (Kollmeier et al. 2015; Smits et al. 2013). Second, open-set formats require experimenter supervision to record participant responses, precluding self-testing or remote administration. Given the increasing focus on telemedicine, accessible home-based testing would offer advantages for both clinical practice and patients by reducing in-person clinical visits (van Wieringen et al. 2021). Equivalent alternatives or complementary testing paradigms for open-set sentence-based testing are needed that accommodate repeated assessment and remote monitoring.

The open-set sentences Leuven Intelligibility Sentence Test (LIST) closely resembles real-world speech and represents an established protocol specifically designed to measure speech perception in hearing impaired individuals and CI users (van Wieringen and Wouters 2008). It features grammatically correct and context-rich sentences that approximate conversational speech (Bench, Kowal, and Bamford 1979; IEEE 1969), giving the LIST high ecological validity. The only factor that reduces its ecological validity is the relatively low speaking rate. However, the test material is constrained by a restricted inventory of 35 available lists of open-set contextual sentences and by the requirement of clinician supervision.

The Matrix test (Hagerman 1982; Kollmeier et al. 2015) offers an alternative methodology using sentences with a fixed-word matrix format. Words are arranged in sentences with a consistent syntactic structure, constraining lexical variability while preserving grammatical integrity. The structure of Matrix sentences inherently reduces memorability even when sentences recur within experiments (Theelen-van den

Hoek, Houben, and Dreschler 2014). However, Matrix sentences have lower ecological validity than the LIST sentences because all sentences follow an identical structure and are not always logically coherent.

The digits-in-noise (DIN) test employs a fundamentally different approach, using three-digit sequences called triplets (Smits et al. 2013) rather than sentences. Although these triplets are less representative of conversational speech, they offer advantages through reduced sensitivity to cognitive capabilities and language proficiency. Triplets have low ecological validity because they do not reflect everyday speech and contain no contextual information. The random triplet combinations inherently minimise memorability compared to coherent sentences.

These three tests differ in their syntactic, lexical, and semantic content, but all were developed to adaptively determine the speech reception threshold (SRT), defined as the signal-to-noise ratio (SNR) where speech recognition performance reaches 50%. All three tests are used with CI users and TH listeners in research and in the clinic. Content learning effects are minimal or negligible for the Matrix and DIN tests (Kollmeier et al. 2015; Smits et al. 2013), making them particularly valuable for repeated testing. However, emerging evidence suggests that perceptual learning effects may occur with the Matrix test (Langerak et al. 2025), a consideration for longitudinal clinical monitoring.

Previous research has demonstrated significant correlations between SRTs of the LIST and DIN test in CI users, with the DIN test explaining 74% of the variance in SRTs of the LIST test (van Wieringen et al. 2021). The DIN test also has been compared to other everyday sentence-in-noise tests. For TH listeners, Smits et al. (2013) reported a Pearson's correlation coefficient of  $r = 0.90$  between the DIN test and the Plomp and Mimpen sentences (Plomp and Mimpen 1979). Kaandorp et al. (2015) compared the DIN test to the VU98 sentence test (Versfeld et al. 2000) across normal hearing (NH) listeners, hearing aid (HA) users, and CI users. While DIN test SRTs explained 86% of the variance in VU98 SRTs overall, the strength of this relationship varied by listener group, with stronger Pearson's correlations observed for NH listeners and HA users ( $r = 0.95$ ) compared to CI users ( $r = 0.56$ ).

The current study was conducted to evaluate whether SRTs obtained with the Matrix and DIN speech materials are comparable to SRTs obtained with the LIST sentences in both CI users and TH listeners. CI users were the primary population of

interest because they show the greatest variability in SRTs and the relationship between the different speech tests is less established in this population. TH listeners were included as a reference group to allow for better comparison with the existing literature, as benchmark studies were typically conducted with TH listeners. In this study, TH listeners were defined as individuals who do not use hearing devices, though they may have age-appropriate hearing loss. Throughout this paper, the term “test” refers to the original speech-in-noise assessment that uses an adaptive procedure. In the current study, however, only the speech stimuli from these tests were used, administered with the method of constant stimuli. This is referred to as the speech material.

To determine the relationship between speech materials, the SRTs obtained with the LIST sentences were regressed against the SRTs of the Matrix sentences or DIN triplets. The regression indicated how closely the SRTs for Matrix sentences or DIN triplets correspond to changes in LIST sentence SRTs, revealing whether these speech materials were equally sensitive to variations in performance with the LIST sentences or whether one was more sensitive than the other. The intercept of the regression represented the baseline difference in SRT between speech materials, which may differ between CI users and TH listeners depending on whether the two groups perceived the relative difficulty of the speech materials similarly. Another aim was to evaluate whether a unified linear regression model could adequately characterise the relationship between the SRTs of LIST sentences and

the other speech materials across both CI users and TH listeners, or whether group-specific models would be required. This analysis addressed whether the conversion between test outcomes depends on hearing status or remains consistent across listener populations. The findings enable audiologists and researchers to make informed decisions about test selection by clarifying how well the Matrix sentences and DIN triplets correspond to the LIST sentences.

## Materials and methods

### Study design and participants

The study included 18 experienced CI users (mean age  $\pm$  SD:  $63 \pm 5$  y) unilaterally implanted with Advanced Bionics devices (Valencia, CA, USA) at the Leiden University Medical Centre (for demographics, see Table 1). Participant selection criteria included monosyllabic consonant-vowel-consonant phoneme scores in quiet of 75% or higher, at least 3 years of CI experience, and age from 18 to 75 years. The 75% CVC threshold was chosen based on the typical post-rehabilitation performance of CI users, which is approximately 80%, ensuring that the sample represents the clinical population (Stronks et al. 2025a). All participants were naive to the material, meaning that they had never heard these particular speech materials before enrolment. Participants were fitted with a Q90 speech processor (Advanced Bionics LLC, Valencia, CA, USA) using their own threshold and maximal comfortable stimulus levels. Front-end processing

**Table 1.** Demographics and characteristics of the cochlear implant users ( $N = 18$ ).

Participant	Age (years)	Gender	CVC (%PH)	Aetiology	CI use (years)	CI model	Processor model
CI01	67	M	90	Unknown, progressive	17	1J	Q90
CI02	61	F	93	Unknown, progressive	23	CII	Q90
CI03	68	M	76	Traumatic*, progressive	5	MS	Q90
CI04	62	M	86	Unknown, progressive	4	MS	Q90
CI05	49	F	95	Unknown, progressive	9	MS	Q90
CI06	66	F	93	DFNA9	11	MS	M90
CI07	59	F	96	Sudden deafness	4	MS	Q90
CI08	62	F	96	Familial*, progressive	4	MS	Q90
CI09	65	M	97	Sudden deafness, Meniere's	4	MS	M90
CI10	59	M	86	Sudden deafness	23	1J	Q90
CI11	75	F	77	Unknown, progressive	6	MS	M90
CI12	62	F	94	Familial*, progressive	5	MS	M90
CI13	71	M	98	Meniere's, progressive	6	MS	M90
CI14	61	F	79	Sudden deafness	19	1J	Q90
CI15	66	M	95	Familial, progressive	23	CII	M90
CI16	61	F	76	Sudden deafness	17	1J	M90
CI17	61	F	87	Unknown, progressive	9	MS	Q90
CI18	64	M	91	DFNA9, progressive	4	MS	Q90
Summary statistics:	62 63 $\pm$ 6	10 F	92 89 $\pm$ 9	NA	8 11 $\pm$ 7	12 MS	11 Q90
Median  mean $\pm$ SD or Sum		8 M				4 1 J	7 M90
						2 CII	

\*Indicates aetiology was presumed but not definitive.

CVC: consonant-vowel-consonant phoneme score in quiet at 65 dB SPL; %PH: percentage of phonemes correct on CVC test; 1J: HiRes 90K HiFocus 1J; CII: Clarion CII HiFocus II; MS: HiRes 90K HiFocus Mid-Scala

strategies (e.g. noise reduction) were deactivated. Any contralateral hearing devices were removed, and the contralateral ear was plugged.

For comparison, a group of 18 age-equivalent TH listeners (mean age  $62 \pm 12$  y) was included. TH listeners did not use hearing devices in daily life and demonstrated pure-tone audiometric thresholds no worse than 30 dB at any tested frequency (125–8000 Hz) compared to the age-specific mean thresholds according to the ISO 7029:2017-06 standard (Kurakata 2023; Michel 2021).

All participants signed informed consent forms. This study received approval from the Medical Ethics Committee Leiden, The Hague, Delft (METC LDD, study number P8.177), and was registered in the Dutch Trial Register of the Central Committee on Research Involving Human Subjects under trial number NL67179.058.18 (<https://onderzoekmetmensen.nl/en/trial/52777>) on October 3, 2022.

### **Test environment**

Speech tests were conducted in a sound-attenuated booth measuring  $3.4 \times 3.2 \times 2.4$  m ( $l \times w \times h$ ). Speech and noise stimuli were presented via a calibrated (Rion NA-28, Rion Co. Ltd., Tokyo, Japan) loudspeaker (KEF, Ci100QS, GP Acoustics, Kent, UK) positioned in front of the participant at ear level at a distance of approximately 1 m.

### **Speech intelligibility testing**

Speech intelligibility was assessed using the material of three validated tests: the LIST sentences (van Wieringen and Wouters 2008), the Dutch/Flemish Matrix sentences (Luts et al. 2014), and the Dutch DIN triplets (Smits et al. 2013). The material of the LIST test consists of 35 lists, each containing 10 sentences of conversational speech of varying lengths and complexity narrated by a Dutch-Flemish female speaker, which can be administered in quiet or noisy backgrounds. It employs sentence scoring, where a sentence is considered correct only when all keywords are accurately repeated. Additionally, the experiment requires continuous experimenter supervision.

The Dutch-Flemish Matrix sentences (Luts et al. 2014) are also voiced by a Dutch-Flemish female speaker. It consists of 13 lists of 20 sentences consisting of semi-random combinations of five words drawn from a closed set of 50, with each sentence composed of a name, verb, quantity, colour, and object. This test is frequently used for CI users and has been validated for TH listeners (Luts et al. 2014).

It was designed as a word-scoring test. When implemented as self-scoring, it functions as a closed-set test where participants select words from a visually presented predefined matrix of options.

The speech material of the Dutch DIN test is voiced by a Dutch male speaker and consists of five unique lists of 24 triplets, expanded to 10 lists through randomisation of each original list (Smits et al. 2013). It employs triplet scoring, where responses are scored as entirely correct or incorrect. Originally developed for telephone-based assessment, the Dutch DIN test can be (self-)administered via computer or mobile device (Kaandorp et al. 2017; Smits et al. 2013). The Dutch DIN test has been validated across hearing-impaired and TH listeners and with CI users (Kaandorp et al. 2015; Smits et al. 2013).

During measurements, participants listened to sentences or triplets and verbally repeated them to the researcher. Word scoring was employed for the Matrix sentences, and triplet and sentence scoring based on keywords were used for the DIN and LIST materials, respectively. Rather than using the adaptive protocols native to each test, the present investigation instead employed the method of constant stimuli to determine psychometric curves (Kingdom and Prins 2016). This approach reduced the required number of presentations to approximately 10 items and enabled dividing the lists for the Matrix and DIN materials, effectively doubling the available list inventory. The measurement protocol was previously described in detail in Stronks et al. (2025b).

Testing was conducted using custom-built software and executed in a MATLAB R2021a programming environment (MathWorks, Inc., Natick, MA, USA). Before testing, participants completed two 10-sentence or 12-triplet practice lists (one presented in quiet and one in noise at +6 dB SNR) to reduce procedural learning (Kollmeier et al. 2015; Smits et al. 2013). For the Matrix sentences, participants had access to a sheet displaying the 50-word matrix during practice. Test conditions were provided in a randomised block design, with one randomised speech material administered on the first testing day, followed by the remaining two speech materials on the second day, also in random order. The sequence of SNRs and list numbers was randomised across participants.

### **Measurement of the SRT**

To determine SRTs, scores were obtained at fixed SNRs, centred on the expected SRT and at SNRs 3, 6, and 9 dB above and below this value. Expected SRTs for

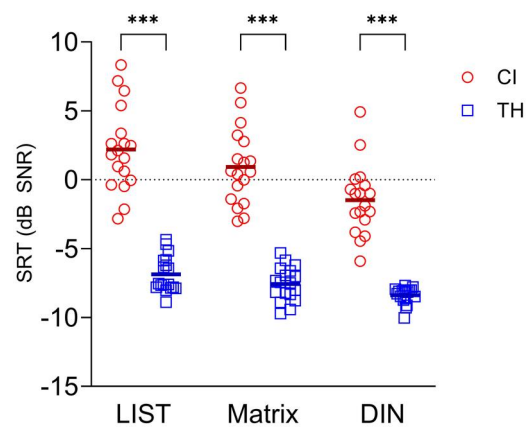
LIST sentences were set at 8 dB SNR for CI users and  $-3$  dB SNR for TH listeners based on previous findings (Claes et al. 2018). For the Matrix sentences, expected SRTs were  $-2$  dB SNR for CI users (Langerak et al. 2024) and  $-4$  dB SNR for TH listeners. The expected SRT for TH listeners was derived by applying a 5 dB SNR age-related adjustment (Claes et al. 2018; van Wieringen and Wouters 2008) to the  $-9$  dB SNR value reported for younger TH populations (Luts et al. 2014). For the DIN triplets, expected SRTs were  $-2$  dB SNR for CI users (Kaandorp et al. 2015) and  $-9$  dB SNR for TH listeners (Kaandorp et al. 2015; Smits et al. 2013). For each speech material, the corresponding long-term speech-shaped noise was presented at a constant level of 60 dBA. Selected SNRs were adjusted in certain cases to ensure that at least seven measurements were available for analysis. Psychometric curves were fitted to the collected data using the reverse Gumbel function (Kingdom and Prins 2016). Fitting was performed with a maximum likelihood procedure using the Palamedes Toolbox version 1.11.11 for MATLAB (Prins and Kingdom 2018).

### Statistical analysis

SRTs were analysed using linear regression to examine the relationships between SRTs obtained with the LIST sentences and those obtained with either the Matrix sentences or DIN triplets across CI and TH groups. Post-hoc analyses employed Šidák's procedure to correct for multiple comparisons. Between-group differences in regression slopes were examined using analysis of covariance, with both slope and intercept parameters tested to characterise group-specific response patterns (Zar 1984). Slopes were compared to a hypothetical 'ideal' slope of 1. Welch's t-test for unequal variances was used to compare SRTs between CI and TH groups. All statistical procedures were conducted using GraphPad Prism version 10.2.3 (GraphPad Software, San Diego, CA, USA), with a significance threshold established at  $p < 0.05$  (two-tailed).

### Results

An independent samples t-test was conducted to examine whether the ages of the CI and TH groups differed significantly. The analysis revealed no significant age difference between the groups,  $t(34) = 0.70$ ,  $p = 0.49$ . The SRTs obtained with the LIST sentences, Matrix sentences, and DIN triplets for CI users and TH listeners are plotted in Figure 1, where lower SRTs represent better performance. For the CI group,



**Figure 1.** Speech reception thresholds (SRTs) across three speech materials for cochlear implant (CI) users and typical hearing (TH) listeners. SRTs (in dB SNR) are shown for the Leuven Intelligibility Sentence Test (LIST) sentences, Matrix sentences, and digits-in-noise (DIN) triplets. Red circles represent individual CI users; blue squares represent individual TH listeners. Horizontal lines indicate group means. Asterisks denote significant difference between groups (\*\*\*)  $p < 0.001$ .

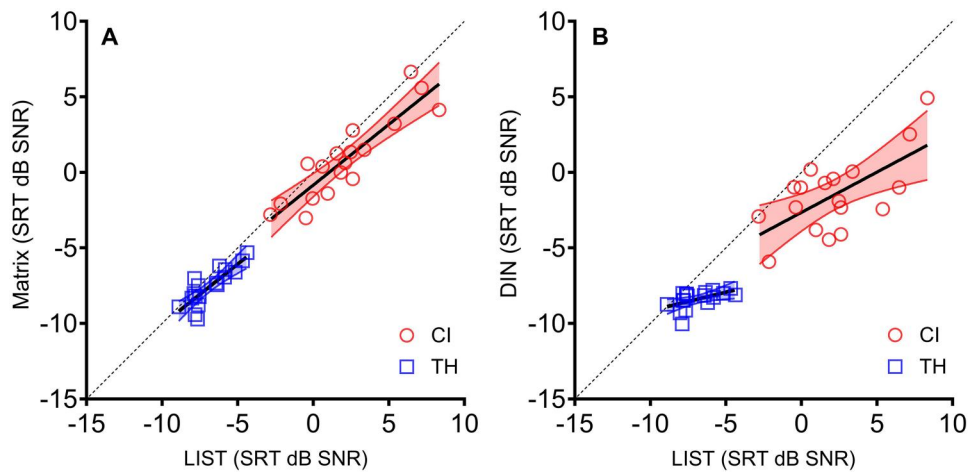
**Table 2.** Descriptive statistics for SRTs using keyword (LIST), word (Matrix), and triplet (DIN) scoring ( $N = 18$  per group).

Group	Speech material	Mean (dB SNR)	SD (dB SNR)	95% CI (dB SNR)
CI	LIST	2.2	3.1	0.7 to 3.7
	Matrix	0.9	2.7	$-0.4$ to 2.3
	DIN	$-1.5$	2.5	$-2.7$ to $-0.2$
TH	LIST	$-6.9$	1.3	$-7.5$ to $-6.2$
	Matrix	$-7.6$	1.2	$-8.2$ to $-7.0$
	DIN	$-8.4$	0.6	$-8.7$ to $-8.3$

CI: cochlear implant; TH: typical hearing; LIST: Leuven Intelligibility Sentence Test; DIN: digits-in-noise

the mean SRTs were: 2.2 dB SNR (SD  $\pm$  3.1 dB) for the LIST sentences, 0.9 dB SNR (SD  $\pm$  2.7 dB) for the Matrix sentences, and  $-1.5$  dB SNR (SD  $\pm$  2.5 dB) for the DIN triplets. SRTs for the TH group were significantly lower than for CI users across all speech materials:  $-6.9$  dB SNR for LIST sentences (SD  $\pm$  1.3 dB,  $p < 0.001$ ),  $-7.6$  dB SNR for Matrix sentences (SD  $\pm$  1.2 dB,  $p < 0.001$ ), and  $-8.4$  dB SNR for DIN triplets (SD  $\pm$  0.6 dB,  $p < 0.001$ ). The descriptive statistics for the SRTs are given in Table 2. Individual psychometric curves from which the SRTs are derived can be found in Stronks et al. (2025c).

The correlations between SRTs obtained with LIST sentences and those obtained with either the Matrix or DIN speech material are presented in Figure 2. Figure 2(A) shows the correlation between LIST sentences and the Matrix sentences, with  $R^2$  values of 0.81 for the CI group and 0.71 for the TH group. The regression slope was significant in the CI group at 0.80 (95% confidence interval [95%CI]: 0.60, 1.01;



**Figure 2.** Linear regression analysis of speech reception thresholds (SRTs) across different speech materials. (A) SRTs in dB SNR for the Leuven Intelligibility Sentence Test (LIST) sentences compared to the Matrix sentences. (B) SRTs for LIST compared to the digits-in-noise (DIN) triplets. Red circles represent cochlear implant (CI) users; blue squares represent typical hearing (TH) listeners. Solid lines show linear regression fits with shaded regions indicating 95% confidence intervals (red for CI users and blue for TH listeners). The dashed line represents the line of unity ( $y = x$ ).

$p < 0.0001$ ) and in the TH group at 0.80 (95%CI: 0.53, 1.07;  $p < 0.0001$ ). The intercepts were  $-0.85$  (95%CI:  $-1.61, -0.07$ ) for the CI group and  $-2.07$  (95%CI:  $-3.97, -0.17$ ) for the TH group.

Regression slopes and intercepts were compared between CI users and TH listeners using analysis of covariance. Post-hoc Šidák's multiple comparisons testing was conducted to examine pairwise differences in regression parameters between groups. For the slopes of the regressed SRTs of the Matrix and LIST sentences, the CI group and TH group did not differ significantly (mean slope difference = 0.0028; 95%CI:  $-0.46, 0.47$ ; adjusted  $p = 0.9999$ ). Similarly, no significant intercept differences were observed between the two groups (mean difference = 1.226; 95%CI:  $-1.08, 3.53$ ; adjusted  $p = 0.4880$ ). A comparison of observed slopes against a hypothetical perfect slope of 1 yielded no significant differences for the CI group ( $p = 0.0629$ ) or the TH group ( $p = 0.1407$ ), indicating that the slopes did not differ significantly from unity.

The results of the regression analysis between SRTs obtained with DIN triplets and LIST sentences are shown in Figure 2(B), with  $R^2$  values of 0.42 for the CI group and 0.28 for the TH group. The model yielded significant slopes for CI users at 0.53 (95%CI: 0.20, 0.87;  $p = 0.0039$ ) and TH listeners at 0.25 (95%CI: 0.04, 0.45;  $p = 0.0225$ ), confirming a significant relationship between SRTs of DIN and LIST materials for both groups. The intercepts were  $-2.65$  (95%CI:  $-3.90, -1.40$ ) for the CI group and  $-6.70$  (95%CI:  $-8.14, -5.27$ ) for the TH group. Post-hoc Šidák's multiple comparisons analysis revealed no significant differences for slopes of DIN and LIST materials between the groups (mean difference = 0.288; 95%CI:  $-0.18, 0.75$ ;

adjusted  $p = 0.3521$ ). However, there were significant intercept differences between the CI and TH groups (mean difference = 4.054; 95%CI: 1.74, 6.36; adjusted  $p = 0.0001$ ), with the TH group demonstrating a more negative intercept ( $-6.704$ ) compared to the CI group ( $-2.65$ ). Tested against an ideal slope of 1, the slopes for both groups differed significantly from unity ( $p = 0.0095$  for CI;  $p < 0.0001$  for TH).

## Discussion

This study presents a comparative analysis of three speech perception materials—LIST sentences, Dutch/Flemish Matrix sentences, and Dutch DIN triplets—administered to both CI users and TH listeners. The primary objective was to determine whether SRTs obtained with the Matrix or DIN materials correlate with those obtained with the LIST sentences, with LIST sentences used as reference for its superior ecological validity and closest representation of real-world speech. Regression analyses revealed significant linear relationships between SRTs of both the Matrix and DIN materials relative to LIST sentence SRTs, although with varying degrees of strength.

The SRTs obtained with the Matrix and LIST sentences demonstrated a strong correlation. The slope for both groups was 0.8, indicating that a 1-dB increase in LIST SRT corresponded to a 0.8 dB increase in Matrix SRT. Neither slopes nor intercepts differed significantly between the CI and TH groups. Moreover, the observed slopes did not significantly differ from unity, suggesting an approximately 1:1 relationship between Matrix and LIST SRTs. These findings demonstrate that SRT

differences measured with the Matrix sentences effectively represent those obtained with everyday sentences for CI users and for TH listeners.

Although both the Matrix and LIST sentences can be administered in closed-set or open-set format, open-set was chosen for this study. The speech material consists of grammatically correct sentences, with rate of speech and sentence context as the primary differences between the two. The speech rate of LIST sentences is reduced to facilitate intelligibility for severely hearing-impaired listeners, including CI users (van Wieringen and Wouters 2008). Unlike LIST sentences, Matrix sentences contain no contextual information because they are semi-random, yet they yield comparable performance patterns across participant groups.

Compared to the correlations between Matrix and LIST sentences, correlations between SRTs from the DIN triplets and LIST sentences were weaker. In contrast to the association between the SRTs of the Matrix and LIST sentences, the regression intercepts for the association of the SRTs between the DIN triplets and LIST sentences differed significantly between the CI and TH groups, reflecting different baseline performance levels for the two populations. The significantly lower intercept for CI users reflects a disproportionate difficulty with the perception of DIN triplets compared to TH listeners. In other words, SRTs obtained with DIN triplets for CI users were lower than would be predicted based on their performance with LIST sentences, and lower than would be predicted from the DIN-LIST relationship observed for TH listeners. This suggests that the digit triplet task presents additional perceptual challenges beyond those captured by sentence recognition materials. The regression slopes for both groups differed significantly from unity, demonstrating a non-proportional relationship between performances on these speech materials. These findings indicate that SRTs from the DIN triplets do not directly correspond to those obtained with everyday sentences.

The lower correlations between SRTs of DIN triplets and SRTs of LIST sentences observed in this study contrast with previous research examining relationships between DIN triplets and sentence-in-noise tests. van Wieringen et al. (2021) reported a stronger relationship between the LIST sentences and the Flemish DIN test, with a slope of 1.098 and  $R^2$  of 0.74 for CI users, compared to a slope of 0.53 in this study. Kaandorp et al. (2015) compared SRTs of the DIN test to those obtained with the VU98 sentences, a Dutch sentence-in-noise test (Versfeld et al. 2000), and reported a shallower regression line for CI users

than for TH participants and hearing aid users. The opposite pattern emerged in the present study, with a shallower regression line for TH listeners compared to CI users. However, when hearing aid users are excluded from the analysis of Kaandorp et al. (2015), the pattern for TH listeners alone is similar to the current findings. Across studies, CI users have demonstrated relatively consistent DIN test SRTs but greater variability in sentence-in-noise tests.

Smits et al. (2013) used an alternative analytical approach, applying linear regression to log-transformed sentence SRTs from the Plomp and Mimpen sentence test (Plomp and Mimpen 1979). This approach generated an exponential growth function describing the relationship between the sentence and DIN test SRTs, with a correlation of  $r=0.90$ . When this transformation method was applied to the current data comparing SRTs for DIN triplets and LIST sentences for TH listeners, the correlation did not strengthen ( $r=0.53$  for both log-transformed and non-transformed LIST SRTs, results not shown), suggesting fundamentally lower correlations between DIN and LIST materials compared to DIN triplets and Plomp and Mimpen sentences.

Several limitations of the current work should be considered. First, constant stimuli were used rather than the adaptive procedure typically employed in clinical protocols for the speech materials. Consequently, the SRTs obtained may differ slightly from those derived using standard clinical methods. Second, two practice lists were used for each speech material to reduce procedural learning effects. This approach is sufficient for DIN triplets (Smits et al. 2013), while LIST sentences have no specified practice requirements (van Wieringen and Wouters 2008). For Matrix sentences, two 10-sentence lists were employed instead of the recommended two 20-sentence lists due to list splitting, which may have resulted in procedural learning during measurements (Kollmeier et al. 2015). However, randomisation of SNR sequences and Matrix sentence lists across participants minimised potential systematic bias from procedural learning effects. Third, the participant sample consisted of unilateral CI users without contralateral assistive devices who were somewhat older than typical research cohorts. While the results apply to this specific population, they are nonetheless representative of the clinical, adult CI population at least in the clinic of this study, and likely in the Netherlands (Stronks et al. 2025a). Fourth, this study focused exclusively on Dutch-speaking populations. Given that the unique syntactic and phonetic characteristics of different languages may influence speech perception test performance, further

research should explore the generalisability of these findings across different languages and linguistic contexts.

The results demonstrate significant correlations between SRTs obtained with Matrix and LIST sentences and between SRTs obtained with DIN triplets and LIST sentences. The steepness of the regression slopes did not differ significantly between CI users and TH listeners. Baseline performance levels, as indicated by intercept values, did not differ between the Matrix and LIST sentences, but differed significantly between DIN and LIST materials. These findings suggest that although both alternatives correlate with performance of the LIST sentences, the Matrix sentences are associated with more consistent relationships across participant groups with different hearing status. With the DIN triplets, the difference in intercepts between hearing modes would complicate comparison of measurements before and after CI implantation.

Repeated speech perception assessment presents a practical challenge. The limited number of LIST sentence lists can lead to content learning effects, creating the need for alternative speech tests. This study shows that the Dutch/Flemish Matrix sentences offer a viable alternative, as Matrix SRTs correlated highly with LIST sentence SRTs for both CI users and TH listeners. For audiologists and researchers requiring repeated assessment, Matrix sentences can be confidently used alongside LIST sentences. By contrast, the DIN triplets showed more limited utility, proving less suitable for comparing pre- and post-implantation outcomes. Beyond addressing the limitations of finite test materials, multiple validated speech perception tests improve clinical flexibility. Expanded materials enable repeated assessment without content learning, while closed-set formats enable remote monitoring. Integrating multiple validated speech perception measures into assessment protocols enhances the quality and accessibility of speech perception evaluation.

## Acknowledgements

We thank the study participants for their time and dedication, and Robin van Deurzen and Paula L. Jansen for their assistance with data collection.

## Author contributions

CRedit: **Bram Knipscheer**: Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing; **Hendrik Christiaan Stronks**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing; **Jeroen Johannes Briaire**: Conceptualization, Formal

analysis, Funding acquisition, Methodology, Project administration, Supervision, Validation, Visualization, Writing – review & editing; **Johan Hubertus Maria Frijns**: Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing.

## Disclosure statement

Johan H.M. Frijns is a member of the European Medical Advisory Board of Advanced Bionics. The other authors have no conflicts of interest to declare.

## Funding

This work was part of the ‘INTENSE’ project within the Crossover Research program funded by the Dutch Research Council (NWO) under grant number 17619. B. Knipscheer is supported by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences and Health, under Grant LSHM20101, to stimulate public-private partnerships. H. Christiaan Stronks, Jeroen J. Briaire, and Johan H. M. Frijns are supported by a non-restrictive research grant from Advanced Bionics LLC (Valencia, CA, USA). Johan H.M. Frijns is a member of the European Medical Advisory Board of Advanced Bionics.

## ORCID

Bram Knipscheer  <http://orcid.org/0000-0001-6341-9405>

Hendrik Christiaan Stronks  <http://orcid.org/0000-0003-1251-8176>

Jeroen Johannes Briaire  <http://orcid.org/0000-0003-4302-817X>

Johan Hubertus Maria Frijns  <http://orcid.org/0000-0002-1180-3314>

## References

- Bench, J., A. Kowal, and J. Bamford. 1979. “The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children.” *British Journal of Audiology* 13 (3): 108–112. <https://doi.org/10.3109/03005367909078884>.
- Billings, C. J., T. M. Olsen, L. Charney, B. M. Madsen, and C. E. Holmes. 2024. “Speech-in-Noise Testing: An Introduction for Audiologists.” *Seminars in Hearing* 45 (1):55–82. <https://doi.org/10.1055/s-0043-1770155>.
- Claes, A. J., P. Van de Heyning, A. Gilles, A. Hofkens-Van den Brandt, V. Van Rompaey, and G. Mertens. 2018. “Impaired Cognitive Functioning in Cochlear Implant Recipients Over the Age of 55 Years: A Cross-Sectional Study Using the Repeatable Battery for the Assessment of Neuropsychological Status for Hearing-Impaired Individuals (RBANS-H).” *Frontiers in Neuroscience* 12: 580. <https://doi.org/10.3389/fnins.2018.00580>.
- Dingemans, J. G., and A. Goedegebure. 2019. “The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests.” *Trends in Hearing* 23:2331216519838672. <https://doi.org/10.1177/2331216519838672>.

- Hagerman, B. 1982. "Sentences for Testing Speech Intelligibility in Noise." *Scandinavian Audiology* 11 (2): 79–87. <https://doi.org/10.3109/01050398209076203>.
- IEEE. 1969. "IEEE Recommended Practice for Speech Quality Measurements." *IEEE Transactions on Audio and Electroacoustics* 17 (3):225–246. <https://doi.org/10.1109/TAU.1969.1162058>.
- Kaandorp, M. W., C. Smits, P. Merkus, S. T. Goverts, and J. M. Festen. 2015. "Assessing Speech Recognition Abilities with Digits in Noise in Cochlear Implant and Hearing Aid Users." *International Journal of Audiology* 54 (1):48–57. <https://doi.org/10.3109/14992027.2014.945623>.
- Kaandorp, M. W., C. Smits, P. Merkus, J. M. Festen, and S. T. Goverts. 2017. "Lexical-Access Ability and Cognitive Predictors of Speech Recognition in Noise in Adult Cochlear Implant Users." *Trends in Hearing* 21:2331216517743887. <https://doi.org/10.1177/2331216517743887>.
- Kingdom, F. A. A., and N. Prins. 2016. *Psychophysics: A Practical Introduction, 2nd edition*. Elsevier Academic Press, London.
- Kollmeier, B., A. Warzybok, S. Hochmuth, M. A. Zokoll, V. Uslar, T. Brand, and K. C. Wagener. 2015. "The Multilingual Matrix Test: Principles, Applications, and Comparison across Languages: A Review." *International Journal of Audiology* 54(sup2):3–16. <https://doi.org/10.3109/14992027.2015.1020971>.
- Kurakata, K. 2023. "Re-Estimated Normal Hearing Threshold Levels for Pure Tones using the Calculation Procedure of ISO 7029:2017." *International Journal of Audiology* 62 (8):720–728. <https://doi.org/10.1080/14992027.2022.2089741>.
- Langerak, N. C., H. C. Stronks, J. J. Briaire, and J. H. M. Frijns. 2024. "The Benefit of Bimodal Hearing and Beamforming for Cochlear Implant Users." *Audiology & Neuro-Otology* 29 (4):297–305. <https://doi.org/10.1159/000536431>.
- Langerak, N. C., H. C. Stronks, J. J. Briaire, and J. H. M. Frijns. 2025. "Learning Effects of the Dutch/Flemish Matrix Test for Bimodal Cochlear Implant Users." *Audiology & Neuro-Otology* :1–11. <https://doi.org/10.1159/000547205>.
- Luts, H., S. Jansen, W. Dreschler, and J. Wouters. 2014. *Development and Normative Data for the Flemish/Dutch Matrix Test*. KU Leuven, Belgium and Academic Medical Center Amsterdam: the Netherlands.
- Michel, O. 2021. "DIN EN ISO 7029:2017-06: Aktuelle DIN-Norm zur Beurteilung der Normalhörigkeit." ["DIN EN ISO 7029:2017-06: The Current DIN Thresholds for Evaluating Normal Hearing." ]*HNO* 69 (12):1014–1018. <https://doi.org/10.1007/s00106-021-01111-3>.
- Plomp, R., and A. M. Mimpen. 1979. "Improving the Reliability of Testing the Speech Reception Threshold for Sentences." *Audiology: official Organ of the International Society of Audiology* 18 (1):43–52. <https://doi.org/10.3109/00206097909072618>.
- Prins, N., and F. A. A. Kingdom. 2018. "Applying the Model-Comparison Approach to Test Specific Research Hypotheses in Psychophysical Research Using the Palamedes Toolbox." *Frontiers in Psychology* 9:1250. <https://doi.org/10.3389/fpsyg.2018.01250>.
- Smits, C., S. T. Goverts, and J. M. Festen. 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>.
- Stronks, H. C., T. S. Arendsen, M. Veenstra, P. B. M. Boermans, J. J. Briaire, and J. H. M. Frijns. 2025a. "Effects of Preoperative Factors on the Learning Curves of Postlingual Cochlear Implant Recipients." *Ear and Hearing* 46 (6):1560–1572. <https://doi.org/10.1097/AUD.0000000000001690>.
- Stronks, H. C., P. L. Jansen, R. van Deurzen, J. J. Briaire, and J. H. M. Frijns. 2025b. "Pupillometry and Perceived Listening Effort for Cochlear Implant Users-A Comparison of Three Speech-In-Noise Tests." *International Journal of Audiology* :1–9. <https://doi.org/10.1080/14992027.2024.2441335>.
- Stronks, H. C., R. van Deurzen, P. L. Jansen, J. J. Briaire, and J. H. M. Frijns. 2025c. "Effect of Speech Material and Scoring Method on Psychometric Curves for Cochlear Implant Users and Typical Hearing Listeners." *Ear and Hearing* 46 (5):1329–1341. <https://doi.org/10.1097/AUD.0000000000001672>.
- Theelen-van den Hoek, F. L., R. Houben, and W. A. Dreschler. 2014. "Investigation into the Applicability and Optimization of the Dutch Matrix Sentence Test for Use with Cochlear Implant Users." *International Journal of Audiology* 53 (11):817–828. <https://doi.org/10.3109/14992027.2014.922223>.
- van Wieringen, A., and J. Wouters. 2008. "LIST and LINT: Sentences and Numbers for Quantifying Speech Understanding in Severely Impaired Listeners for Flanders and the Netherlands." *International Journal of Audiology* 47 (6):348–355. <https://doi.org/10.1080/14992020801895144>.
- van Wieringen, A., S. Magits, T. Francart, and J. Wouters. 2021. "Home-Based Speech Perception Monitoring for Clinical Use With Cochlear Implant Users." *Frontiers in Neuroscience* 15:773427. <https://doi.org/10.3389/fnins.2021.773427>.
- Versfeld, N. J., L. Daalder, J. M. Festen, and T. Houtgast. 2000. "Method for the Selection of Sentence Materials for Efficient Measurement of the Speech Reception Threshold." *The Journal of the Acoustical Society of America* 107 (3): 1671–1684. <https://doi.org/10.1121/1.428451>.
- Wagener, K. C., and T. Brand. 2005. "Sentence Intelligibility in Noise for Listeners with Normal Hearing and Hearing Impairment: Influence of Measurement Procedure and Masking Parameters." *International Journal of Audiology* 44 (3):144–156. <https://doi.org/10.1080/14992020500057517>.
- Winn, M. B. 2016. "Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants." *Trends in Hearing* 20 1–17. <https://doi.org/10.1177/2331216516669723>.
- Yund, E. W., and D. L. Woods. 2010. "Content and Procedural Learning in Repeated Sentence Tests of Speech Perception." *Ear and Hearing* 31 (6):769–778. <https://doi.org/10.1097/AUD.0b013e3181e68e4a>.
- Zar, J. 1984. "Comparing Simple Linear Regression Equations." In *Biostatistical Analysis*, 2nd ed. Prentice-Hall, Englewood Cliffs.