Pupillometry and perceived listening effort for cochlear implant users—a comparison of three speech-in-noise tests

Stronks, Hendrik Christiaan; Jansen, Paula Louisa; van Deurzen, Robin; Briaire, Jeroen Johannes; Frijns, Johan Hubertus Maria

DOI
10.1080/14992027.2024.2441335

Publication date
2025

Document Version
Final published version

Published in
International Journal of Audiology

International Journal of Audiology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iija20

Pupillometry and perceived listening effort for cochlear implant users—a comparison of three speech-in-noise tests

Hendrik Christiaan Stronks, Paula Louisa Jansen, Robin van Deurzen, Jeroen Johannes Briaire & Johan Hubertus Maria Frijns

To cite this article: Hendrik Christiaan Stronks, Paula Louisa Jansen, Robin van Deurzen, Jeroen Johannes Briaire & Johan Hubertus Maria Frijns (20 Jan 2025): Pupillometry and perceived listening effort for cochlear implant users—a comparison of three speech-in-noise tests, International Journal of Audiology, DOI: 10.1080/14992027.2024.2441335

To link to this article: https://doi.org/10.1080/14992027.2024.2441335

INTERNATIONAL JOURNAL OF AUDIOLOGY
https://doi.org/10.1080/14992027.2024.2441335

ORIGINAL ARTICLE

OPEN ACCESS

Pupillometry and perceived listening effort for cochlear implant users—a comparison of three speech-in-noise tests

Hendrik Christiaan Stronks[a,b], Paula Louisa Jansen[a], Robin van Deurzen[a], Jeroen Johannes Briaire[a] and Johan Hubertus Maria Frijns[a,b,c]

[a]Department of Otorhinolaryngology and Head & Neck Surgery, Leiden University Medical Center, Leiden, Netherlands; [b]Leiden Institute for Brain and Cognition, Leiden, Netherlands; [c]Department of Bioelectronics, Delft University of Technology, Delft, Netherlands

ABSTRACT

Objective: Measuring listening effort using pupillometry is challenging in cochlear implant (CI) users. We assess three validated speech tests (Matrix, LIST, and DIN) to identify the optimal speech material for measuring peak-pupil-dilation (PPD) in CI users as a function of signal-to-noise ratio (SNR).

Design: Speech tests were administered in quiet and two noisy conditions, namely at the speech recognition threshold (0 dB re SRT), i.e. the SNR where speech intelligibility (SI) was 50%, and at a more favourable SNR of +6 dB re SRT. PPDs and subjective ratings of effort were obtained.

Study sample: Eighteen unilaterally implanted CI users.

Results: LIST sentences revealed significantly different PPDs between +6 and 0 dB re SRT and DIN triplets between quiet and +6 dB re SRT. PPDs obtained with the Matrix test were independent of SNR and yielded large PPDs and high subjective ratings even in quiet.

Conclusions: PPD is a sensitive measure for listening effort when processing LIST sentences near 0 dB re SRT and when processing DIN triplets at more favourable listening conditions around +6 dB re SRT. PPDs obtained with the Matrix test were insensitive to SNR, likely because it is demanding for CI users even in quiet.

ARTICLE HISTORY
Received 1 March 2024
Revised 2 December 2024
Accepted 7 December 2024

KEYWORDS
Sensorineural hearing loss; cochlear implants; pupillometry; speech in noise; subjective measures

Introduction

Cochlear implants (CIs) are the first-line treatment for severe-to-profound sensorineural hearing loss. A CI bypasses the degenerated hair cells in the cochlea and directly stimulates the auditory nerve electrically (Naples and Ruckenstein 2020). CIs perform very well in quiet, but background noise severely degrades speech recognition for CI users and substantially more so than for typical-hearing listeners (Cullington and Zeng 2008). The sensitivity to noise makes listening cognitively demanding for CI users, and they suffer more from listening fatigue, especially in adverse listening conditions (McGarrigle et al. 2014; Perreau et al. 2017). For these reasons, it is crucial to assess hearing performance not in terms of speech recognition alone but to include measures of listening effort as well to arrive at a more comprehensive evaluation of listening capabilities (Humes 1999). Listening effort is the mental exertion required to understand speech (McGarrigle et al. 2014) and reflects the total amount of cognitive resources needed for a particular listening task (Zekveld, Koelewijn, and Kramer 2018).

To assess the outcomes of CI in noise and evaluate the effectiveness of speech enhancement paradigms, such as noise reduction algorithms, a suitable measure of listening effort should depend on the signal-to-noise ratio (SNR). The SNR is a convenient variable for adapting task difficulty in speech recognition testing, and higher (more favourable) SNRs are expected to be associated with a decrease in listening effort. A multitude of outcome measures of listening effort have been developed that capture different aspects of listening effort occurring at different points in time (Shields et al. 2023).

Subjective measures of listening effort, including Likert-scaled ratings, reflect late aspects of listening effort (Shields et al. 2023) and associate significantly and robustly with SNR and thus speech intelligibility in typical hearing listeners, hearing aid users and CI users alike (Stronks, Apperloo, et al. 2021; Stronks et al. 2024; Zekveld, Kramer, and Festen 2010). Their drawback is their susceptibility to observer bias (Moore and Picou 2018), and subjective ratings are considered less reliable than objective measures of listening effort (Giuliani, Brown, and Wu 2021). Behavioural measures, including auditory response-time measurements and dual-task paradigms, capture events of intermediate latency and have shown some potential in CI users (Pals et al. 2015).

Objective measures of listening effort capture the earliest components of effortful listening. EEG has shown the potential to quantify listening effort (Burle et al. 2015), yet its use with CI users is complicated because of the electrical artefacts arising from the electrodes (Paul et al. 2021; Zhang et al. 2010). fMRI shows promise (Burle et al. 2015) but is usually not compatible with CI because of the magnetisable components (Saksida et al. 2021). Other physiological measures, such as heart rate and skin conductance suffer from neither drawback but have yielded mixed results (Giuliani, Brown, and Wu 2021). A

CONTACT Hendrik Christiaan Stronks h.c.stronks@lumc.nl Department of Otorhinolaryngology and Head & Neck Surgery, Leiden University Medical Center, PO Box 9600, Leiden, 2300 RC, Netherlands

Supplemental data for this article can be accessed online at https://doi.org/10.1080/14992027.2024.2441335.

well-established measure of listening effort is pupillometry (Winn, Edwards, and Litovsky 2015; Zekveld, Koelewijn, and Kramer 2018; Zekveld, Kramer, and Festen 2010). It is based on the observation that the pupil dilates when listening, and this response increases under more challenging listening conditions. The most straightforward and accepted way to measure this is using peak pupil dilation (PPD).

Pupillometry is compatible with CI and does not suffer from electrical artefacts (Winn and Moore 2018). SNR in CI users affects pupil response latency and release of effort (Dingemanse and Goedegebure 2022). However, unlike typical hearing listeners and hard-of-hearing people, the more conventional PPD measure is insensitive to widely varying SNRs in CI users. Our group arrived at this conclusion using Matrix sentences (Stronks, Apperloo, et al. 2021; Stronks et al. 2024), whereas others used meaningful sentences in noise, namely the Versfeld (Dingemanse and Goedegebure 2022) and HINT sentences (Zhang et al. 2023). While the absence of SNR effects on the PPD may be due to the characteristics of listening with a CI, experimental conditions cannot be ruled out. For instance, due to the syntactic structure of randomly drawn words, the Matrix sentences do not allow postdiction ('filling in the gaps'), which we hypothesised to play a role in the lack of SNR effect on the PPD (Stronks et al. 2024). The Versfeld sentences allow for postdiction but are uttered by a swift-talking male speaker, which may not be suitable for all CI users. Lastly, the study from Zhang et al. (2023) tested a range of fixed SNRs; because not all CI users perform equally well in noise, this inadvertently caused different participants to operate at varying performance levels and, hence, different task difficulties at one particular SNR.

In this study, we investigated the dependency of PPDs on SNR in CI users using the Matrix, LIST, and DIN tests. Subjective ratings were collected to capture the downstream consequences of listening effort. The LIST material was developed specifically for use in CI users and consists of meaningful sentences including sufficient context for postdiction (Van Wieringen and Wouters 2008). We hypothesised that the LIST sentences are more suitable than the Matrix material to measure PPD as a function of SNR in CI users. The DIN test consists of random triplets of numbers without context (Smits, Goverts, and Festen 2013) that do not allow for postdiction. However, because of its simplicity and small number of words per target, it is cognitively the least demanding task. Under the assumption that CI users operate at near-ceiling effort even in quiet (Dingemanse and Goedegebure 2022), this characteristic was hypothesised to decrease cognitive load and make the DIN test a promising candidate for measuring listening effort in CI users.

Materials and methods

Study design and participants

This prospective crossover study included 18 experienced, unilateral CI users implanted with Advanced Bionics devices (Valencia, CA, USA). They were recruited from the Leiden University Medical Centre (Table 1) and selected based on their CVC phoneme score in quiet (>75%), experience with their CI (>3 years), and age (adults ≤ 75 years). After obtaining informed consent, contralateral hearing devices were removed, and the ear was plugged. The participants were fitted with a research speech processor (Q90, Advanced-Bionics LLC, Valencia, CA, USA) using their threshold and maximal comfortable stimulus levels. Any front-end processing strategies (e.g. noise reduction algorithms) were deactivated. This study was approved by the medical ethical committee (IRB) of Leiden, Den Haag, Delft (METC LDD) under study number P8.177. It was included in the Dutch Trial Register of the Central Committee on Research Involving Human Subjects (CCMO) under trial number NL67179.058.18 (https://onderzoekmetmensen.nl/en/trial/52777) on 3 October 2022. This research adhered to the tenets of Helsinki (World Medical Association 2013).

Table 1. Demographics and characteristics of the study participants (*N* = 18).

ID	Age	Sex	CVC	Etiology	Years CI	CI	Processor
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
CI12	62	F	94	Familial?, progressive	5	MS	M90
CI11	75	F	77	Unknown, progressive	6	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
Mean	63	10 F	89		11	12 MS	11 Q90
SD	6	8 M	8		7	4 1j	7 M90
Median	62		92		8	2 CII	

CVC: consonant-vowel-consonant phoneme score in quiet at 65 dB SPL; DFNA9: 1j: HiRes 90K HiFocus 1j; CII: Clarion CII HiFocus II; MS: HiRes 90K HiFocus Mid-Scala.

Test environment

Participants were seated in the middle of a sound-attenuated booth measuring 3.4 × 3.2 × 2.4 m (l × w × h). Speech and noise were presented by the same loudspeaker in front of the participant at ear level (KEF, Ci100QS, GP Acoustics, Kent, UK) at ~1 m distance. Sound stimuli were calibrated with a sound level metre (Rion NA-28, Rion Co. Ltd., Tokyo, Japan).

Speech recognition testing

Speech recognition was assessed using three validated tests: the Dutch/Flemish Matrix test (Luts et al. 2014), LIST (Van Wieringen and Wouters 2008), and DIN (Smits, Goverts, and Festen 2013). Participants listened to each sentence (or triplet) and verbally repeated it to the experimenter. Guessing was allowed, but no feedback was provided during testing. Each set of lists was preceded by two practice lists, one in quiet and the other in noise (+6 dB SNR). In the case of the Matrix test, a sheet with the 50-word Matrix was available to the participant for the duration of the practice lists. The respective long-term speech-shaped LTSS noise was used for each test and presented at a constant level of 60 dBA. Noise preceded and followed the target by 2 s.

Each speech test has different syntactic, lexical, and semantic content. Still, all were developed to adaptively determine the speech reception threshold, defined as the SNR where speech recognition performance is 50%. Every test had its specific adaptive protocol defined when it was initially introduced.

The Matrix test consists of 13 lists of 20 meaningless sentences with semi-random combinations of five words drawn from a closed set of 50, such that each sentence is grammatically correct and composed of a name, verb, quantity, colour, and object. Each of the five words is scored, and the SNR of the following sentence is determined by a custom algorithm, such as the OLSA (Brand and Kollmeier 2002).

The LIST test consists of 35 sets of 10 meaningful sentences of varying lengths and difficulty. It usually adopts sentence scoring based on keywords and a fixed step size of 2 dB to determine the SRT. Because the sentences are coherent and meaningful, postdiction is possible based on the context.

The DIN test consists of five lists of 24 triplets, which was extended to ten lists by randomising each of the original ones. It was initially benchmarked using triplet scoring and an adaptive protocol with a fixed step size. Because the numbers are random, postdiction is hardly possible. The number of possibilities of the remaining digits in a triplet decreases when one is correctly perceived due to a process of elimination because each digit is unique. However, we did not provide the participants with this information.

We adapted these procedures to enhance comparability across tests. We maintained the word scoring for the Matrix test. Still, we deployed keyword scoring for the LIST and digit scoring for the DIN test to make them more lenient for CI users, who generally cannot reach a 100% word score even in quiet.

Speech tests were conducted using custom-built software and were executed in a MATLAB R2021a programming environment (MathWorks, Inc., Natick, MA, USA). Test conditions were randomised within and across participants.

We determined the SRT, defined as the SNR where speech recognition was 50%, as a reference to establish three fixed SNRs for the pupillometry. By testing SNRs based on SRTs, instead of using a fixed set of SNRs in all participants, we ensured that the task was demanding yet manageable for each participant individually. We did not determine the SRT adaptively, but we constructed psychometric curves to assess the SRT to allow for better comparisons between tests.

Construction of psychometric curves

Psychometric curves were determined using the method of constant stimuli by measuring the speech recognition score at seven fixed SNRs relative to the expected SRT based on the literature, namely at SRT and at SNRs 3, 6, and 9 dB above and below that. Each SNR was evaluated with 10 Matrix and LIST sentences or 12 DIN triplets. The data were fitted with an inverse log-Weibull (Gumbel) function (Gilchrist, Jerwood, and Ismaiel 2005) in a MATLAB programming environment (R2021a) using the least squares method with the "lsqcurvefit" function.

Pupillometry

Pupillometry was performed at three fixed SNRs that were determined for each participant individually using their SRT, namely SRT, SRT + 6 dB, and quiet (SNR = ∞). SNRs were chosen relative to SRT to match task difficulty across participants. SRT was the most challenging condition and expected to yield the largest PPD (Stronks et al. 2024). Speech tests and SNRs were evaluated in a randomised block design.

Pupil diameter was recorded using eye-tracking glasses (ETG 2.6, SensoriMotor Instruments, Teltow, Germany) with a spatial resolution of 1280 × 960 pixels and a sample rate of 120 Hz. The room was illuminated by a dimmable LED strip mounted on the room's walls above eye level. The luminance was set for each participant individually by measuring six different light intensities between 2 and 200 lx measured at eye level (Sonel LXP-2, Świdnica, Poland), such that the pupil was approximately in the middle of its dynamic range, expectedly to be between 4.5 and 5.0 mm (Watson and Yellott 2012).

The experimental paradigm for pupillometry and signal post-processing was based on Stronks, Apperloo, et al. (2021) and Stronks et al. (2024). Each target sentence/triplet was preceded by 2 s quiet and 2 s of LTSS noise at 60 dBA. The target was followed by 2 s of noise and 3 s of quiet, whereafter a probe tone was played at ~60 dBA that signalled the participant to repeat the target (Figure 1) verbally.

Eye blinks were defined by a pupil diameter of <2 mm and eliminated by linear interpolation from 10 samples before the start of the blink to 10 samples after. Pupil sizes >8 mm were treated as artefacts because they fall outside the range of normal pupil diameters (Watson and Yellott 2012) and were eliminated via the same interpolation method. The signal was then low-pass filtered using an 8th-order Butterworth filter with a cut-off frequency of 5 Hz to reduce remaining artefacts. The average diameter during the last second of the baseline LTSS noise recording was defined as the baseline diameter to determine the pupil response.

Sweeps were discarded after deblinking when ≥50% of baseline samples were interpolated, when 25% or more of the post-baseline recording was interpolated, and when an uninterrupted stretch of >1 s of interpolation was present. The remaining accepted traces were baseline-corrected, and a final rejection step was introduced to eliminate traces with gross residual artefacts with an overall amplitude (the maximum minus minimum of the entire sweep) larger than the overall amplitude + 2·*SD* of the ensemble average. The remaining traces were ensemble averaged.

If the total number of included waveforms from one of two eyes was <5, it was discarded from all three SNRs in the respective speech test, and the PPD was determined in the remaining eye to maintain proper analysis of the SNR within each speech test. If both eyes did not meet the criterion, the SNR was dropped for that participant, but the remaining two SNRs were maintained, yielding missing data. If two eyes were available in

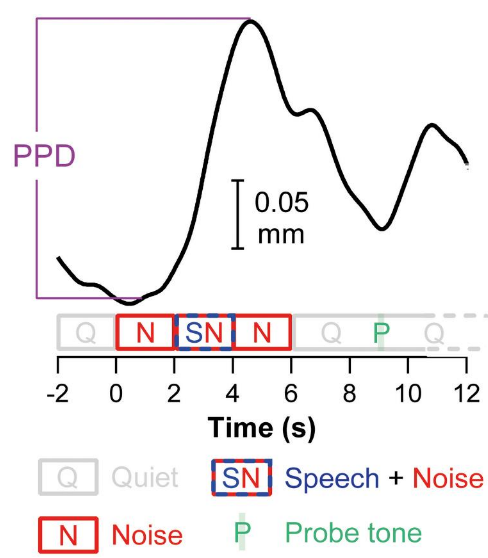

Figure 1. Sample waveform of the average pupil response in the CI group performing the DIN test at 0 dB re SRT. The stimulus paradigm is shown below the waveform, including a 2-s quiet period (Q), 2 s of noise (N), the stimulus and noise (SN), followed by another 2 s of N and Q and a probe tone (P) signalling the participant to repeat the stimulus. The Peak pupil dilation (PPD) was defined as the difference between peak and baseline.

all conditions, the PPDs from both eyes were averaged with equal weight. This procedure resulted in one missing SNR (both eyes lost) for the Matrix and the DIN test, both from participant S12 at SNR = SRT. The mean number of included traces per eye was 17 (Matrix, LIST) and 22 (DIN, having 24 targets instead of 20).

Subjective listening effort and performance ratings

After completing a list at a particular SNR, participants were asked to subjectively rate their listening effort on a nine-point Likert scale from 1 (no effort) to 9 (very effortful) (Stronks, Apperloo, et al. 2021; Stronks et al. 2024). Participants were encouraged to ignore procedural task difficulty and fatigue when rating listening effort.

Statistical analysis

Statistical significance testing was performed with linear mixed models fitted with the restricted maximum likelihood (REML) procedure using IBM SPSS Statistics for Windows 29.0 (released 2022, IBM Corp. Armonk, NY, USA). Factors of interest, for instance, the type of speech test (Matrix, LIST or (DIN), SNR (0 dB and +6 dB re SRT, or quiet), and their interaction were included as repeated fixed factors. The covariance matrices of the repeated and random variables were set at scaled identity. This matrix structure is, by default, used in SPSS when a single random variable is present. For the repeated factors, the choice was based on an iterative process applied for each outcome variable (PPD, effort rating, performance). To this end, different covariance structures suitable to the kind of longitudinal data under analysis were compared using REML estimation (Wolfinger 1993), including "diagonal," "scaled identity," "compound symmetry" and its correlation and heterogeneous variants, and "unstructured" and its correlations variant. Because of the small sample size, we relied on the Bayesian information criterion (Gurka 2006), and degrees of freedom were determined according to Kenward-Roger, which is recommended for small samples (Kenward and Roger 1997). *Post-hoc* multiple comparisons significance tests were performed on the estimated marginal (EM) means between pairs of conditions using Šidák's correction.

Correlation analyses were conducted to determine the coefficient of determination (r^2) using standard linear regression with the least squares method and subsequent F-tests on the slope with Prism 9.5.1 for Windows (GraphPad Software, San Diego, CA, USA). RM ANOVA was used to compare the SRTs between speech tests also with GraphPad Prism.

Results

Pupillometry

The SNRs were determined relative to the SRT. The SRT was obtained individually for each participant and each speech test. The averaged SRT ± SD across participants for the Matrix test was 1 ± 3 dB SNR, for LIST 0 ± 3 dB SNR, and for the DIN test −5 ± 1 dB SNR. Testing was performed at SRT re 0 dB, at a more favourable SNR of +6 dB re SRT where noise was decreased by 6 dB, and in quiet where the speech level was the same as in the other two conditions. A Geisser-Greenhouse-corrected RM ANOVA [$F(2,30) = 98$, $p < 0.001$] showed that the mean SRT of the DIN test was significantly lower than that of the Matrix (Tukey's corrected multiple comparisons test; $p < 0.001$) and LIST test ($p < 0.001$). The Matrix and LIST tests did not differ significantly in this respect ($p = 0.084$).

Population-averaged waveforms of the pupil diameter obtained in quiet (in purple), +6 dB re SRT (blue), and at SRT (green) are shown in Figure 2. For all speech tests, pupil responses increased at less favourable SNRs. This effect was, however, small for the Matrix test. In the case of the DIN test, the difference between +6 dB re SRT and SRT appeared negligible.

Figure 3 shows the averaged PPDs (A), ratings of listening effort (B), and correct scores (C). The effects of the type of speech material and SNR were tested for significance by constructing a full factorial linear mixed model for each of the three outcome measures with the following structure:

$$Y = SNR + Test + (Test*SNR)|participant \qquad (1)$$

where Y is the outcome measure (PPD, effort rating, or performance), SNR and $Test$ are repeated fixed factors representing the signal-to-noise ratio and the type of speech test, respectively, and $|participant$ is the random participant factor. Significant main effects of SNR [$F(2,134) = 12.799$, $p < 0.001$] and the interaction term [$F(4,134) = 4.062$, $p = 0.004$] on the PPD were found. The main effect of the speech test on PPD was not significant [$F(2,134) = 0.119$, $p = 0.888$]. Based on the significant main interaction between SNR and type of speech test, a Šidák's corrected *post-hoc* multiple comparisons test was performed to individually compare the three SNRs for each speech test (Table 2). No significant difference in PPD between any of the SNR pairs was found for the Matrix test. Significant PPDs were observed between 0 dB re SRT and the other two conditions for the LIST test. By contrast, the DIN test showed significant differences between the quiet and the other conditions. A second *post-hoc* test was performed to compare the SNRs between speech tests (Table 2). In quiet, the Matrix test was associated with significantly larger PPDs than the LIST and DIN tests, whereas the LIST and DIN tests did not differ significantly. No significant differences in PPD were found between speech tests at less favourable SNRs.

Ratings of listening effort

LMM analysis of the ratings of listening effort revealed significant main effects of SNR [$F(2,136) = 117.505$, $p < 0.001$], type of speech test [$F(2,136) = 9.164$, $p < 0.001$], and the interaction factor [$F(4,136) = 6.346$, $p < 0.001$]. Based on the interaction effect, a Šidák's corrected *post-hoc* multiple comparisons test was performed to compare the Likert ratings at the three SNRs for each speech test (Table 3). Effort ratings differed significantly between all the tested SNR pairs for all speech tests, except between +6 and 0 dB re SRT for the Matrix test. The ratings were consistently highest at more challenging SNRs, such that they followed 0 dB re SRT > +6 dB re SRT > quiet. *Post-hoc* testing of the ratings between speech tests for each SNR (Table 3) showed that the Matrix test was rated significantly and substantially higher (more subjective effort) than the LIST test in quiet and at +6 dB re SRT. The Matrix test also showed higher ratings than the DIN test in quiet. No significant differences were observed between LIST and DIN at any SNR, and at 0 dB re SRT, ratings were not significantly different between any of the tests.

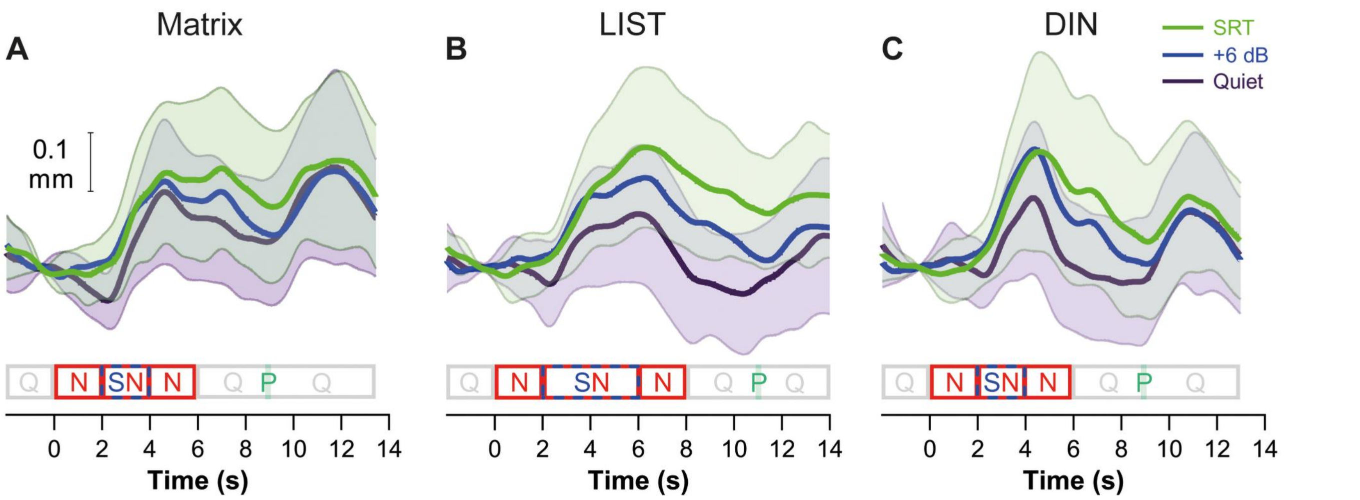

Figure 2. Average waveforms obtained with the different speech tests from CI users in quiet (purple ± *SD*), +6 dB re SRT (blue, nor error margins for clarity), and at SRT (green ± *SD*). The stimulus paradigm is indicated above the time axis; Q: 2 s of quiet; N: 2 s of noise; SN: signal (sentence or triplet) plus noise; N: 2 s of noise; Q: 3 s of quiet; P: short probe tone; Q: variable duration of quiet where the participant verbally repeated what was heard. N was 2 s of quiet when no noise was presented (quiet situation). Error margins: *SD*.

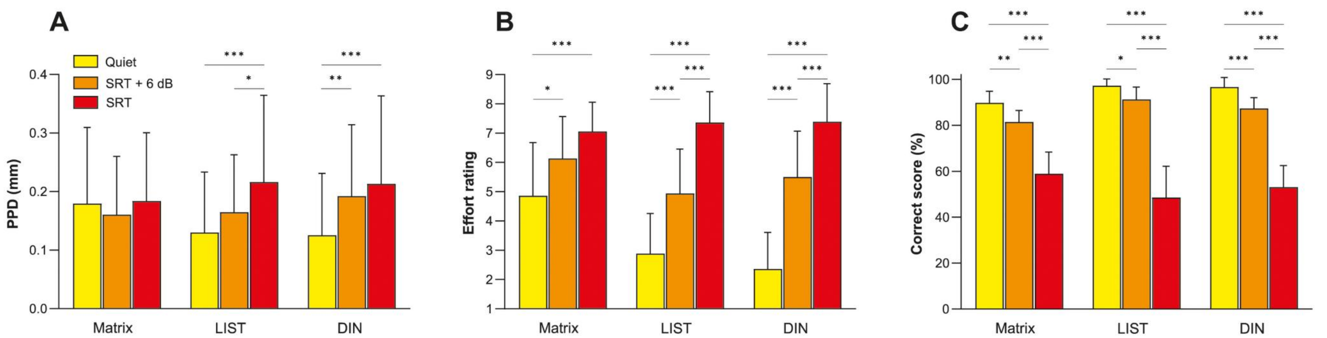

Figure 3. Overview of the peak pupil response (PPD, A,B), ratings of listening effort (C,D), and correct scores (E,F) obtained with three speech tests and listening difficulties in typical hearing participants (TH, left) and cochlear implant users (CI, right). Yellow bars: quiet; orange: SRT + 6 dB; red: SRT. Error bars: *SD*. Significant differences between SNRs within speech tests are indicated by asterisks; *$p < 0.05$, **$p < 0.01$; ***$p < 0.001$. For differences in SNR pairs between speech tests (see main text).

Table 2. *Post-hoc*, Šidák's corrected multiple comparisons tests of PPD.

Speech test	SNR re SRT (I)	SNR re SRT (II)	I–II (mm)	$CI_{95\%}$	*p*	SNR	Speech test (I)	Speech test (II)	I–II (mm)	$CI_{95\%}$	*p*
Matrix	0 dB	Quiet	0.00	−0.05–0.05	1.000	0 dB	Matrix	LIST	−0.04	−0.08–0.01	0.205
	0 dB	+6 dB	0.02	−0.03–0.07	0.665	re SRT	Matrix	DIN	−0.03	−0.08–0.02	0.376
	+6 dB	Quiet	−0.02	−0.07–0.03	0.688		LIST	DIN	0.01	−0.04–0.05	0.984
LIST	0 dB	Quiet	0.09*	0.04–0.13	<0.001	+6 dB	Matrix	LIST	0.00	−0.05–0.04	0.994
	0 dB	+6 dB	0.05*	0.00–0.10	0.027	re SRT	Matrix	DIN	−0.03	−0.08–0.02	0.280
	+6 dB	Quiet	0.04	−0.01–0.08	0.204		LIST	DIN	−0.03	−0.07–0.02	0.411
DIN	0 dB	Quiet	0.08*	0.04–0.13	<0.001	Quiet	Matrix	LIST	0.05*	0.00–0.10	0.033
	0 dB	+6 dB	0.02	−0.03–0.07	0.749		Matrix	DIN	0.05*	0.01–0.10	0.017
	+6 dB	Quiet	0.07*	0.02–0.11	0.002		LIST	DIN	0.01	−0.04–0.05	0.993

$df = 134$, $SE = 0.02$.
*Difference significant.

Correct scores

The LMM with correct scores as outcome measure showed that they depended significantly on SNR [$F(2,136) = 468.222$, $p < 0.001$] and the interaction between SNR and the type of speech test [$F(4,136) = 10.472$, $p < 0.001$]. There was no significant main effect of the kind of speech test [$F(2,136) = 1.777$, $p = 0.173$]. Šidák's corrected *post-hoc* testing showed that speech scores significantly differed between every pair of SNRs for every speech test (Table 4). According to expectation, scores were better at higher SNRs, following 0 dB re SRT < +6 dB re SRT < quiet. *Post-hoc* tests between speech tests showed that the Matrix test significantly differed from the other two tests at favourable SNRs, whereas the DIN and LIST tests did not differ significantly at any SNR (Table 4).

Correlation of the PPD between speech tests

The dependence of PPD on SNR differed between speech tests. To correlate the SNR sensitivity between speech tests, we calculated a slope measure *via*:

$$PPD_{slope} = \frac{PPD_{0dB\ re\ SRT} - PPD_{+6dB\ re\ SRT}}{6} \quad (2)$$

Table 3. *Post-hoc* multiple comparisons tests of effort rating.

Speech test	SNR re SRT (I)	SNR re SRT (II)	I–II (Likert)	$CI_{95\%}$	p	SNR	Speech test (I)	Speech test (II)	I–II (Likert)	$CI_{95\%}$	p
Matrix	0 dB	Quiet	2.2*	1.1–3.3	<0.001	0 dB re SRT	Matrix	LIST	−0.3	−1.4–0.8	0.761
	0 dB	+6 dB	0.9	−0.2–2.0	0.114		Matrix	DIN	−0.3	−1.4–0.7	0.733
	+6 dB	Quiet	1.3*	0.2–2.3	0.013		LIST	DIN	0.0	−1.1–1.0	1.000
LIST	0 dB	Quiet	4.5*	3.4–5.5	<0.001	+6 dB re SRT	Matrix	LIST	1.2*	0.1–2.3	0.023
	0 dB	+6 dB	2.4*	1.4–3.5	<0.001		Matrix	DIN	0.6	−0.4–1.7	0.386
	+6 dB	Quiet	2.1*	1.0–3.1	<0.001		LIST	DIN	−0.6	−1.6–0.5	0.507
DIN	0 dB	Quiet	5.0*	4.0–6.1	<0.001	Quiet	Matrix	LIST	2.0*	0.9–3.0	<0.001
	0 dB	+6 dB	1.9*	0.8–3.0	<0.001		Matrix	DIN	2.5*	1.4–3.6	<0.001
	+6 dB	Quiet	3.1*	2.1–4.2	<0.001		LIST	DIN	0.5	−0.5–1.6	0.550

df = 136, *SE* = 0.4.
*Difference significant.

Table 4. *Post-hoc* multiple comparisons tests of correct scores.

Speech test	SNR re SRT (I)	SNR re SRT (II)	I–II (%)	$CI_{95\%}$	p	SNR	Speech test (I)	Speech test (II)	I–II (mm)	$CI_{95\%}$	p
Matrix	0 dB	Quiet	−31*	−37 to −25	<0.001	0 dB re SRT	Matrix	LIST	10*	4 to 16	<0.001
	0 dB	+6 dB	−23*	−28 to −17	<0.001		Matrix	DIN	6	0 to 12	0.057
	+6 dB	Quiet	−8*	−14 to −2	0.003		LIST	DIN	−5	−11 to 1	0.187
LIST	0 dB	Quiet	−49*	−55 to −43	<0.001	+6 dB re SRT	Matrix	LIST	−10*	−16 to −4	<0.001
	0 dB	+6 dB	−43*	−49 to −37	<0.001		Matrix	DIN	−6	−12 to 0	0.054
	+6 dB	Quiet	−6*	−12 to 0	0.048		LIST	DIN	4	−2 to 10	0.300
DIN	0 dB	Quiet	−44*	−50 to −38	<0.001	Quiet	Matrix	LIST	−8*	−13 to −2	0.008
	0 dB	+6 dB	−34*	−40 to −28	<0.001		Matrix	DIN	−7*	−13 to −1	0.017
	+6 dB	Quiet	−9*	−15 to −3	<0.001		LIST	DIN	1	−5 to 7	0.994

df = 136, *SE* = 2.5.
*Difference significant.

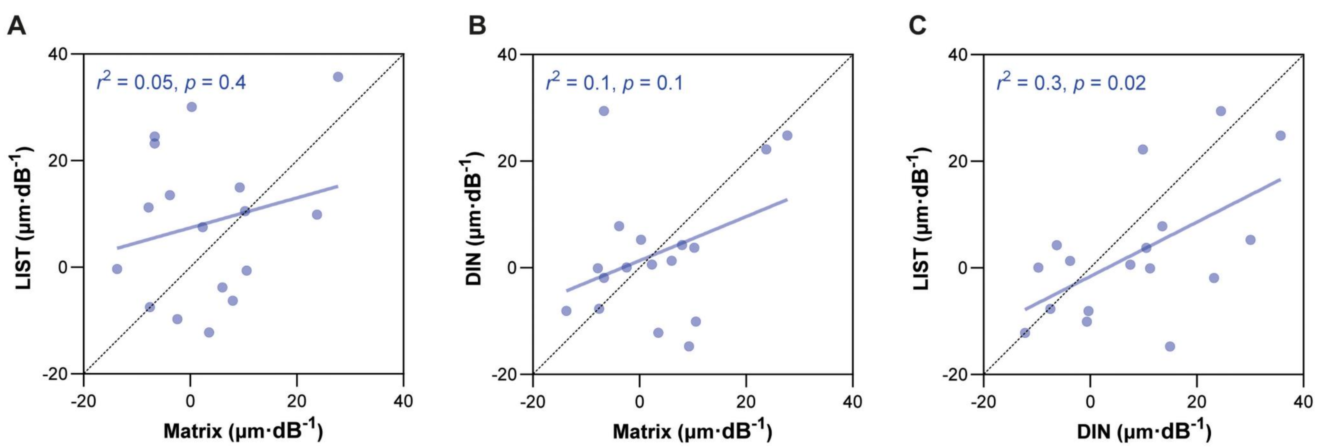

Figure 4. Correlation of PPD sensitivity to change in SNR between the LIST and Matrix test (A), LIST and DIN (B), and Matrix and DIN (C). Solid blue line: correlation obtained with linear regression; dashed black line: hypothetical correlation with identical slopes.

where PPD_{slope} is the sensitivity of the PPD to SNR, $PPD_{0dB\ re\ SRT}$ and $PPD_{+6dB\ re\ SRT}$ the PPD obtained at 0 dB re SRT and +6 dB re SRT, respectively, and 6 is the SNR difference in dB. A positive PPD_{slope} reflected a larger pupil diameter increase in the more challenging condition (Figure 4). The correlation between the PPD change obtained with the LIST and Matrix test (Figure 4(A)) and DIN and Matrix test (Figure 4(B)) was not significant [$r^2 = 0.05$, $F(1,15) = 0.752$, $p = 0.399$ and $r^2 = 0.133$, $F(1,15) = 2.303$, $p = 0.150$, respectively], whereas it was significant between the LIST and DIN test [Figure 4(C), $r^2 = 0.333$, $F(1,15) = 7.477$, $p = 0.0154$].

Correlation between objective and subjective measures of listening effort

The same procedure was applied to look for correlations between the change in PPD and the change in the rating of listening effort, and between change in PPD and correct score (Figure 5). Not in any of the speech tests a significant correlation between the change in PPD and change in effort rating (red symbols and regression lines) was found [Matrix test (Figure 5(A)): $r^2 = 0.037$, $F(1,15) = 0.572$, $p = 0.461$, LIST test (Figure 5(B)): $r^2 = 0.066$, $F(1,16) = 1.127$, $p = 0.304$; DIN test (Figure 5(C)): $r^2 = 0.121$, $F(1,15) = 2.072$, $p = 0.171$] nor between change in PPD and change in correct score [blue symbols and regression lines; Matrix test (Figure 5(A)): $r^2 < 0.001$, $F(1,15) = 0.006$, $p = 0.942$, LIST test (Figure 5(B)): $r^2 < 0.001$, $F(1,16) = 0.003$, $p = 0.958$; DIN test (Figure 5(C)): $r^2 = 0.015$, $F(1,15) = 0.227$, $p = 0.641$].

Discussion

Performance with CI is often assessed by measures of speech recognition, but other outcomes, notably listening effort, can be

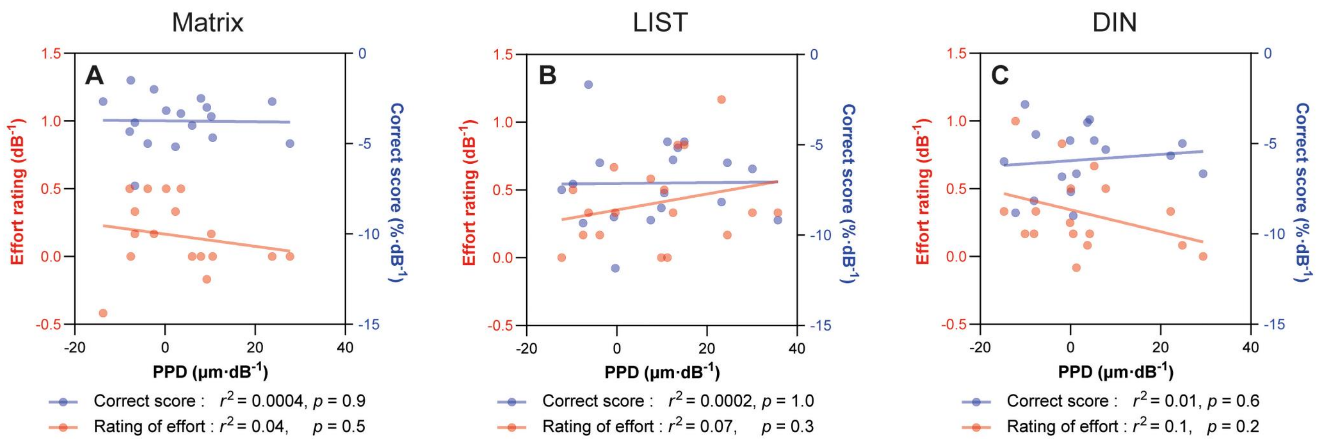

Figure 5. Sensitivities of effort ratings and correct scores to SNR correlated with the sensitivity of the PPD for the Matrix (A), LIST (B), and DIN test (C). Red symbols: effort ratings; blue symbols: correct ratings; solid lines: correlation obtained with linear regression.

equally important for the end user. In this research, we have compared three widely used speech tests with differing lexical, syntactic, and semantic characteristics to identify the most promising alternative to measure listening effort as a function of SNR in CI users with pupillometry. This work followed up on earlier findings showing that the PPD was not dependent on a wide range of SNRs when the Matrix test was adopted in CI users (Stronks, Apperloo, et al. 2021; Stronks et al. 2024), whereas significant effects were found in typical hearing participants. We corroborate these findings here by showing that PPDs obtained with the Matrix test were not significantly dependent on SNR for the Matrix test. By contrast, PPDs obtained with the LIST and DIN test significantly increased at more challenging SNRs. PPDs obtained with LIST sentences differed most low SNRs (around the SRT), whereas those obtained with DIN triplets differed at higher SNRs where correct scores approached the plateau.

By correlating the difference between the PPD obtained at SRT and SRT +6 dB across speech tests, we show a significant relation between the LIST and DIN tests but not between the Matrix test and the other two, suggesting that the Matrix test is processed cognitively differently in CI users. This was surprising because the speech corpora of the LIST and Matrix tests have more apparent similarities than LIST and DIN. Notably, the Matrix and LIST speech material consisted of sentences uttered by a Flemish female speaker, whereas the DIN material consisted of digits spoken by a Dutch male.

One of the reasons for the apparent SNR-insensitivity of the PPD for the Matrix test is its apparent difficulty. PPDs and ratings of effort were significantly higher in quiet for the Matrix test, and correct scores were lower than those of the other tests. Higher ratings and lower scores than the LIST test were also observed at +6dB re SRT. A reason for this can be the high rate of speech of the Matrix sentences (∼5 words in 2 s). The DIN test features only 3 digits in the same amount of time, whereas the LIST sentences are voiced at even slower rates (2.5 syllables/second) (Van Wieringen and Wouters 2008).

Additionally, the DIN test puts little demand on linguistic and cognitive skills. It relies less on working memory than other speech tests because of the small number of items per stimulus (Kaandorp et al. 2017). Because of these considerations, a simplified Matrix test has been developed for populations, such as older adults and children for whom working memory or lexical abilities may be limiting (Prang et al. 2021; Willberg et al. 2020). The name and verb have been removed from the sentence, thus leaving only the numeral, adjective, and object. A similar approach is followed in the coordinate response measure (CRM), where only 2 keywords are scored (Moore 1981). Assessing the merits of these simplified Matrix tests and CRM for pupillometric purposes for CI users will be relevant. Additionally, we cannot rule out that regular Matrix sentences developed in other languages may yield PPDs responsive to SNR for CI users, nor that they may prove helpful in different experimental paradigms, such as dual response tasks.

Unlike the other two tests, the LIST sentences allowed for postdiction based on context. Context-rich sentences require less vigilance or cognitive control and are associated with smaller pupil responses (Winn and Teece 2021). Assuming that listening is highly demanding for CI users and that they exert near-ceiling effort even in quiet (Dingemanse and Goedegebure 2022), cognitively less demanding tasks (here: LIST and DIN) can decrease cognitive load and increase the “dynamic range” of PPDs available to CI users.

At 0 dB re SRT, correct scores obtained with the Matrix test (but not with LIST or DIN) were substantially higher than the expected 50% (59%, Figure 3(C)), likely due to learning effects (Stronks, Briaire, and Frijns 2020). This may have influenced the results at this SNR and point to another possible limitation of the Matrix test for assessing listening effort.

Changes in PPD in response to changing SNRs around 0 dB re SRT did not correlate significantly with changes observed in the ratings of effort, nor with the correct scores in any of the three speech tests, corroborating earlier findings (Stronks et al. 2024) and supporting the notion that physiological measures and subjective ratings of listening effort reflect different components of cognition (Francis et al. 2016).

In terms of detecting changes in listening effort, the SNR range around SRT is of particular interest because the psychometric curve is typically the steepest (Strasburger 2001), and the pupil response largest (Ohlenforst et al. 2017). The LIST sentences revealed a significant effect of SNR on the PPD around SRT, which is hence preferred over the DIN test. The latter is better suited when PPDs are obtained at more favourable SNRs.

For both the DIN and LIST tests, it is relevant to know the sensitivity of the tests to smaller SNR increments than those used here. For instance, many speech enhancement strategies and fitting configurations for CIs improve the SRT by only a few dB SNR (e.g. Goldsworthy et al. 2014; Plant et al. 2016; Stronks, Briaire, and Frijns 2020; Stronks, Tops, et al. 2021).

We used scientifically and clinically widely accepted speech tests, but the downside is that they differed in many respects, complicating the interpretation of the results. Hence, it is of interest to identify the characteristics that set the DIN and LIST speech materials apart from the Matrix test. The most obvious differences were the lexical (vocabular) and syntactic (structural) content (Verma and Vasan Srinivasan 2019), yet other subtle differences that may have affected cognitive processing also existed. For instance, the LIST stimuli were of longer duration than the others and the lack of semantic context in the Matrix and DIN stimuli can have influenced task engagement. The fact that the LIST sentences were voiced by a Flemish female, whereas a Dutch male uttered the DIN triplets may have also influenced the processing of the material.

A sustained pupil dilation after the PPD could be observed at 0 dB re SRT that differed somewhat between tests (Figure 2(A)), pointing to differences in the release of effort (Winn 2016). We captured this effect by determining the difference between the PPD and the following negative peak in the waveform. Still, this procedure yielded highly variable outcomes without any correlation to SNR (results not shown). Previous attempts to capture temporal aspects of the pupil response also failed, including peak latencies or time-integrated responses (Stronks, Apperloo, et al. 2021). Generalised additive mixed models may prove helpful to further investigate the temporal effects of SNR (Abramowitz, Goupell, and DeRoy Milvae 2024).

Conclusion

We conclude that LIST sentences are suitable for measuring listening effort with pupillometry in CI users at SNRs around SRT, whereas DIN is the preferred test at more favourable SNRs. We tentatively conclude that cognitively undemanding speech material is most optimal for measuring PPDs in CI users. The Dutch-Flemish Matrix test yielded PPDs that were unresponsive to changes in SNR and it proved to be a challenging listening test even in quiet. Additional research is needed to verify whether Matrix sentences in other languages similarly yield PPDs unresponsive to SNR and whether simplified variants of the Matrix test are better suited for this purpose.

Acknowledgements

We are grateful to the study participants for their time and dedication. We thank Nicolas Furnon (Advanced Bionics, European Research Center, Hannover, Germany) for technical support.

Disclosure statement

The authors H. Christiaan Stronks, Johan H. M. Frijns, and Jeroen J. Briaire 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.

Funding

This work was partly funded by the Dutch Research Council (NWO) within the Crossover Research program under the "INTENSE" project (grant #17619) and co-funded by Advanced Bionics.

ORCID

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

Abramowitz, J. C., M. J. Goupell, and K. DeRoy Milvae. 2024. "Cochlear-Implant Simulated Signal Degradation Exacerbates Listening Effort in Older Listeners." *Ear and Hearing* 45 (2): 441–450. https://doi.org/10.1097/AUD.0000000000001440.

Brand, T., and B. Kollmeier. 2002. "Efficient Adaptive Procedures for Threshold and Concurrent Slope Estimates for Psychophysics and Speech Intelligibility Tests." *The Journal of the Acoustical Society of America* 111 (6): 2801–2810. https://doi.org/10.1121/1.1479152.

Burle, B., L. Spieser, C. Roger, L. Casini, T. Hasbroucq, and F. Vidal. 2015. "Spatial and Temporal Resolutions of EEG: Is It Really Black and White? A Scalp Current Density View." *International Journal of Psychophysiology* 97 (3): 210–220. https://doi.org/10.1016/j.ijpsycho.2015.05.004.

Cullington, H. E., and F. G. Zeng. 2008. "Speech Recognition with Varying Numbers and Types of Competing Talkers by Normal-Hearing, Cochlear-Implant, and Implant Simulation Subjects." *The Journal of the Acoustical Society of America* 123 (1): 450–461. https://doi.org/10.1121/1.2805617.

Dingemanse, G., and A. Goedegebure. 2022. "Listening Effort in Cochlear Implant Users: The Effect of Speech Intelligibility, Noise Reduction Processing, and Working Memory Capacity on the Pupil Dilation Response." *Journal of Speech, Language, and Hearing Research* 65 (1): 392–404. https://doi.org/10.1044/2021_JSLHR-21-00230.

Francis, A. L., M. K. MacPherson, B. Chandrasekaran, and A. M. Alvar. 2016. "Autonomic Nervous System Responses During Perception of Masked Speech May Reflect Constructs Other than Subjective Listening Effort." *Frontiers in Psychology* 7: 263. https://doi.org/10.3389/fpsyg.2016.00263.

Gilchrist, J. M., D. Jerwood, and H. S. Ismaiel. 2005. "Comparing and Unifying Slope Estimates Across Psychometric Function Models." *Perception & Psychophysics* 67 (7): 1289–1303. https://doi.org/10.3758/bf03193560.

Giuliani, N. P., C. J. Brown, and Y. H. Wu. 2021. "Comparisons of the Sensitivity and Reliability of Multiple Measures of Listening Effort." *Ear and Hearing* 42 (2): 465–474. https://doi.org/10.1097/AUD.0000000000000950.

Goldsworthy, R. L., L. A. Delhorne, J. G. Desloge, and L. D. Braida. 2014. "Two-Microphone Spatial Filtering Provides Speech Reception Benefits for Cochlear Implant Users in Difficult Acoustic Environments." *The Journal of the Acoustical Society of America* 136 (2): 867–876. https://doi.org/10.1121/1.4887453.

Gurka, M. J. 2006. "Selecting the Best Linear Mixed Model Under REML." *The American Statistician* 60 (1): 19–26. https://doi.org/10.1198/000313006X90396.

Humes, L. E. 1999. "Dimensions of Hearing Aid Outcome." *Journal of the American Academy of Audiology* 10 (1): 26–39.

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.

Kenward, M. G., and J. H. Roger. 1997. "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood." *Biometrics* 53 (3): 983–997. https://doi.org/10.2307/2533558.

Luts, H., S. Jansen, W. Dreschler, and J. Wouter. 2014. *Development and Normative Data for the Flemish/Dutch Matrix Test.* Amsterdam: Katholieke Universiteit Leuven, Belgium and Academic Medical Center.

McGarrigle, R., K. J. Munro, P. Dawes, A. J. Stewart, D. R. Moore, J. G. Barry, and S. Amitay. 2014. "Listening Effort and Fatigue: What Exactly Are We Measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'White Paper'." *International Journal of Audiology* 53 (7): 433–440. https://doi.org/10.3109/14992027.2014.890296.

Moore, T. J. 1981. "Voice Communication Jamming Research." In *AGARD Conference Proceedings 331: Aural Communication in Aviation.* Neuilly-Sur-Seine, France, 6: 1–2.

Moore, T. M., and E. M. Picou. 2018. "A Potential Bias in Subjective Ratings of Mental Effort." *Journal of Speech, Language, and Hearing Research* 61 (9): 2405–2421. https://doi.org/10.1044/2018_JSLHR-H-17-0451.

Naples, J. G., and M. J. Ruckenstein. 2020. "Cochlear Implant." *Otolaryngologic Clinics of North America* 53 (1): 87–102. https://doi.org/10.1016/j.otc.2019.09.004.

Ohlenforst, B., A. A. Zekveld, T. Lunner, D. Wendt, G. Naylor, Y. Wang, N. J. Versfeld, and S. E. Kramer. 2017. "Impact of Stimulus-Related Factors and Hearing Impairment on Listening Effort as Indicated by Pupil Dilation." *Hearing Research* 351: 68–79. https://doi.org/10.1016/j.heares.2017.05.012.

Pals, C., A. Sarampalis, H. van Rijn, and D. Başkent. 2015. "Validation of a Simple Response-Time Measure of Listening Effort." *The Journal of the Acoustical Society of America* 138 (3): EL187–EL192. https://doi.org/10.1121/1.4929614.

Paul, B. T., J. Chen, T. Le, V. Lin, and A. Dimitrijevic. 2021. "Cortical Alpha Oscillations in Cochlear Implant Users Reflect Subjective Listening Effort During Speech-in-Noise Perception." *PLOS One* 16 (7): e0254162. https://doi.org/10.1371/journal.pone.0254162.

Perreau, A. E., Y.-H. Wu, B. Tatge, D. Irwin, and D. Corts. 2017. "Listening Effort Measured in Adults with Normal Hearing and Cochlear Implants." *Journal of the American Academy of Audiology* 28 (8): 685–697. https://doi.org/10.3766/jaaa.16014.

Plant, K., R. van Hoesel, H. McDermott, P. Dawson, and R. Cowan. 2016. "Influence of Contralateral Acoustic Hearing on Adult Bimodal Outcomes After Cochlear Implantation." *International Journal of Audiology* 55 (8): 472–482. https://doi.org/10.1080/14992027.2016.1178857.

Prang, I., M. Parodi, C. Coudert, S. Legoff, M. Exter, M. Buschermöhle, F. Denoyelle, and N. Loundon. 2021. "The Simplified French Matrix. A Tool for Evaluation of Speech Intelligibility in Noise." *European Annals of Otorhinolaryngology, Head and Neck Diseases* 138 (4): 253–256. https://doi.org/10.1016/j.anorl.2020.12.003.

Saksida, A., S. Ghiselli, S. Bembich, A. Scorpecci, S. Giannantonio, A. Resca, P. Marsella, and E. Orzan. 2021. "Interdisciplinary Approaches to the Study of Listening Effort in Young Children with Cochlear Implants." *Audiology Research* 12 (1): 1–9. https://doi.org/10.3390/audiolres12010001.

Shields, C., M. Sladen, I. A. Bruce, K. Kluk, and J. Nichani. 2023. "Exploring the Correlations Between Measures of Listening Effort in Adults and Children: A Systematic Review with Narrative Synthesis." *Trends in Hearing* 27: 23312165221137116. https://doi.org/10.1177/23312165221137116.

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.

Strasburger, H. 2001. "Converting Between Measures of Slope of the Psychometric Function." *Perception & Psychophysics* 63 (8): 1348–1355. https://doi.org/10.3758/bf03194547.

Stronks, H. C., A. L. Tops, K. W. Quach, J. J. Briaire, and J. H. M. Frijns. 2024. "Listening Effort Measured with Pupillometry in Cochlear Implant Users Depends on Sound Level, But Not on the Signal to Noise Ratio When Using the Matrix Test." *Ear and Hearing* 45 (6): 1461–1473. https://doi.org/10.1097/AUD.0000000000001529.

Stronks, H. C., A. L. Tops, P. Hehrmann, J. J. Briaire, and J. H. M. Frijns. 2021. "Personalizing Transient Noise Reduction Algorithm Settings for Cochlear Implant Users." *Ear and Hearing* 42 (6): 1602–1614. https://doi.org/10.1097/AUD.0000000000001048.

Stronks, H. C., E. Apperloo, R. Koning, J. J. Briaire, and J. H. M. Frijns. 2021. "Softvoice Improves Speech Recognition and Reduces Listening Effort in Cochlear Implant Users." *Ear and Hearing* 42 (2): 381–392. https://doi.org/10.1097/AUD.0000000000000928.

Stronks, H. C., J. J. Briaire, and J. H. M. Frijns. 2020. "The Temporal Fine Structure of Background Noise Determines the Benefit of Bimodal Hearing for Recognizing Speech." *Journal of the Association for Research in Otolaryngology* 21 (6): 527–544. https://doi.org/10.1007/s10162-020-00772-1.

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.

Verma, G., and B. Vasan Srinivasan. 2019. "A Lexical, Syntactic, and Semantic Perspective for Understanding Style in Text." *arXiv e-prints* arXiv: 1909.08349.

Watson, A. B., and J. I. Yellott. 2012. "A Unified Formula for Light-Adapted Pupil Size." *Journal of Vision* 12 (10): 12. https://doi.org/10.1167/12.10.12.

Willberg, T., K. Kärtevä, M. Zokoll, M. Buschermöhle, V. Sivonen, A. Aarnisalo, H. Löppönen, B. Kollmeier, and A. Dietz. 2020. "The Finnish Simplified Matrix Sentence Test for the Assessment of Speech Intelligibility in the Elderly." *International Journal of Audiology* 59 (10): 763–771. https://doi.org/10.1080/14992027.2020.1741704.

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: 2331216516669723. https://doi.org/10.1177/2331216516669723.

Winn, M. B., and A. N. Moore. 2018. "Pupillometry Reveals that Context Benefit in Speech Perception Can Be Disrupted by Later-Occurring Sounds, Especially in Listeners with Cochlear Implants." *Trends in Hearing* 22: 2331216518808962. https://doi.org/10.1177/2331216518808962.

Winn, M. B., and K. H. Teece. 2021. "Listening Effort Is Not the Same as Speech Intelligibility Score." *Trends in Hearing* 25: 23312165211027688. https://doi.org/10.1177/23312165211027688.

Winn, M. B., J. R. Edwards, and R. Y. Litovsky. 2015. "The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation." *Ear and Hearing* 36 (4): e153–e165. https://doi.org/10.1097/AUD.0000000000000145.

Wolfinger, R. 1993. "Covariance Structure Selection in General Mixed Models." *Communications in Statistics-Simulation and Computation* 22 (4): 1079–1106. https://doi.org/10.1080/03610919308813143.

World Medical Association. 2013. "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects." *JAMA* 310 (20): 2191–2194. https://doi.org/10.1001/jama.2013.281053.

Zekveld, A. A., S. E. Kramer, and J. M. Festen. 2010. "Pupil Response as an Indication of Effortful Listening: The Influence of Sentence Intelligibility." *Ear and Hearing* 31 (4): 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251.

Zekveld, A. A., T. Koelewijn, and S. E. Kramer. 2018. "The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge." *Trends in Hearing* 22: 2331216518777174. https://doi.org/10.1177/2331216518777174.

Zhang, F., J. Anderson, R. Samy, and L. Houston. 2010. "The Adaptive Pattern of the Late Auditory Evoked Potential Elicited by Repeated Stimuli in Cochlear Implant Users." *International Journal of Audiology* 49 (4): 277–285. https://doi.org/10.3109/14992020903321759.

Zhang, Y., M. A. Callejón-Leblic, A. M. Picazo-Reina, S. Blanco-Trejo, F. Patou, and S. Sánchez-Gómez. 2023. "Impact of SNR, Peripheral Auditory Sensitivity, and Central Cognitive Profile on the Psychometric Relation Between Pupillary Response and Speech Performance in CI Users." *Frontiers in Neuroscience* 17: 1307777. https://doi.org/10.3389/fnins.2023.1307777.