Check for updates

Impact of SNR, peripheral auditory sensitivity, and central cognitive profile on the psychometric relation between pupillary response and speech performance in CI users

Yue Zhang[1]*, M. Amparo Callejón-Leblic[2,3,4], Ana M. Picazo-Reina[3], Sergio Blanco-Trejo[3], François Patou[5] and Serafín Sánchez-Gómez[3]

[1]Department of Research and Technology, Oticon Medical, Vallauris, France, [2]Oticon Medical, Madrid, Spain, [3]ENT Department, Virgen Macarena University Hospital, Seville, Spain, [4]Biomedical Engineering Group, University of Sevillel, Sevillel, Spain, [5]Department of Research and Technology, Oticon Medical, Smørum, Denmark

Despite substantial technical advances and wider clinical use, cochlear implant (CI) users continue to report high and elevated listening effort especially under challenging noisy conditions. Among all the objective measures to quantify listening effort, pupillometry is one of the most widely used and robust physiological measures. Previous studies with normally hearing (NH) and hearing-impaired (HI) listeners have shown that the relation between speech performance in noise and listening effort (as measured by peak pupil dilation) is not linear and exhibits an inverted-U shape. However, it is unclear whether the same psychometric relation exists in CI users, and whether individual differences in auditory sensitivity and central cognitive capacity affect this relation. Therefore, we recruited 17 post-lingually deaf CI adults to perform speech-in-noise tasks from 0 to 20 dB SNR with a 4 dB step size. Simultaneously, their pupillary responses and self-reported subjective effort were recorded. To characterize top-down and bottom-up individual variabilities, a spectro-temporal modulation task and a set of cognitive abilities were measured. Clinical word recognition in quiet and Quality of Life (QoL) were also collected. Results showed that at a group level, an inverted-U shape psychometric curve between task difficulty (SNR) and peak pupil dilation (PPD) was not observed. Individual shape of the psychometric curve was significantly associated with some individual factors: CI users with higher clinical word and speech-in-noise recognition showed a quadratic decrease of PPD over increasing SNRs; CI users with better non-verbal intelligence and lower QoL showed smaller average PPD. To summarize, individual differences in CI users had a significant impact on the psychometric relation between pupillary response and task difficulty, hence affecting the interpretation of pupillary response as listening effort (or engagement) at different task difficulty levels. Future research and clinical applications should further characterize the possible effects of individual factors (such as motivation or engagement) in modulating CI users' occurrence of 'tipping point' on their psychometric functions, and develop an individualized method for reliably quantifying listening effort using pupillometry.

1 Introduction

Cochlear implantation is a standard care for individuals with severe-to-profound sensorineural hearing loss who obtain limited benefit from conventional hearing aids. Typically, cochlear implant (CI) recipients show good levels of speech communication in quiet, but the high inter-subject variability observed in outcome measures remains a challenge in both clinical and research fields (Lazard et al., 2012; Holden et al., 2013; Boisvert et al., 2020; Goudey et al., 2021). Specifically, CI users find it difficult to listen in complex daily-life situations and noisy environments. Such variability may be due to many factors including, but not limited to, the age of implantation, duration of deafness, duration of CI use, the limited spectral and temporal resolution of the signal, quality of the electrode-neural interface, and the narrow electric dynamic range of the output signals codified at the stimulation electrodes (Blamey et al., 2012; Holden et al., 2013; Dorman and Gifford, 2017; Boisvert et al., 2020). CI users often complain of the high listening effort required in those situations, even when speech communication is successful (Hughes and Galvin, 2013; Winn et al., 2015; Pichora-Fuller et al., 2016; Winn, 2016; Rapport et al., 2020; Dingemanse and Goedegebure, 2022; Winn and Teece, 2022). Eventually, this sustained and unresolved high effort status associated with hearing could have negative consequence on social communication and cognitive health for CI users (Nachtegaal et al., 2009; Pichora-Fuller et al., 2015; Herrmann and Johnsrude, 2020; Shukla et al., 2020).

Although speech recognition is a conventional metric of CI performance, it does not capture the listening effort experienced by CI users when they expend cognitive resources to achieve a certain level of speech performance. Quantifying listening effort provides an additional layer of information to complement and inform the efficiency of speech intelligibility measures. Different methods to quantify listening effort have been proposed, including physiological, subjective, and behavioral methods. Together, these methods reveal different underlying domains of listening effort (Strand et al., 2018; Shields et al., 2023). Specifically, pupillometry is one of the most widely used physiological methods. Pupillometry has been proved to be a robust technique in registering changes in cognitive effort (Kramer et al., 1997; Zekveld and Kramer, 2014; Zekveld et al., 2018). Different studies have demonstrated that pupillometry measures can capture the intensity of effort, and it has been already documented to be sensitive to speech intelligibility, spectral resolution, linguistic complexity, attention, concurrent cognitive demands, background noise, etc. (Koelewijn et al., 2012, 2014; Besser et al., 2013; Winn et al., 2015; Ohlenforst et al., 2017; Zekveld et al., 2019; Bönitz et al., 2021; Micula et al., 2021; Zhang et al., 2021). Pupillometry also holds great potential to be adapted into the clinical setting due to its non-invasive nature and its robustness against electrical CI artifacts.

Benefits and limitations of CI on speech recognition have been extensively researched. Relatively fewer studies have measured listening effort in CI users, even though the findings from the limited studies all suggest the importance of including listening

effort measures when assessing CI outcomes. One of the difficulties preventing more listening effort studies (specifically using pupillometry), both in research and clinics, is our lack of understanding of the psychometric relation between CI users' speech performance and listening effort measured by pupillary response. As shown by previous studies (Ohlenforst et al., 2017, 2018; Wendt et al., 2017), the relation between speech performance in noise and listening effort (as measured by peak pupil response) is not linear in normal hearing (NH) and hearing-impaired (HI) participants: listeners show bigger peak pupil response (PPD), one of the most widely used pupillary marker for listening effort, as the sentence recognition in noise task gets more difficult, until they reach a certain "tipping point" and disengage from the task as it gets overly difficult. From there onwards, listeners show smaller PPD as the task increases in difficulty. On a group level, the psychometric relation between task difficulty and pupillary response showed an inverted U-shape: in high SNR regions, higher SNR relates to lower PPD; and in low SNR regions higher SNR relates to bigger PPD. Many studies have used this psychometric relation as the basis to interpret changes in pupil response as the effectiveness of intervening strategies in reducing listening effort for NH and HI listeners (Winn, 2016; Ohlenforst et al., 2018). However, no study has yet validated this relation for CI users across a similarly wide range of listening conditions (i.e., 9 SNR steps with speech recognition score ranging from 0 to 100%). It is highly likely that CI users do not have similar hearing and physiological status as NH and HI listeners, thus CI users' psychometric relation between listening task and pupillary response might not be the same as in NH and HI listeners (Hughes and Galvin, 2013; Wang et al., 2016; Winn, 2016; Perreau et al., 2017; Perea Pérez et al., 2023). Therefore, without such validation with CI users on this piece of evidence, it is difficult to establish enough confidence to use pupillometry to interpret listening effort and assess efficacy of new CI interventions, such as new signal processing strategies, noise reduction and beamforming algorithms, among others.

Furthermore, past studies have investigated the relation between CI speech outcomes and individual auditory and cognitive factors. But it is unclear whether and how individual differences in auditory sensitivity and central cognitive resources affect listening effort and the psychometric relation between listening conditions and pupillary responses in CI users. It is unlikely that all CI users would bear the same psychometric relations, considering the big variabilities in CI outcomes. Therefore, it is important to identify and quantify the effects of individual factors that contribute to the variabilities in the psychometric relation. Multiple studies have shown significant correlation between spectral-temporal acuity with speech performance, suggesting that bottom-up auditory processing ability is important to explain CI variabilities (Aronoff and Landsberger, 2013; Lawler et al., 2017; DiNino and Arenberg, 2018). In addition to biological and audiological factors, individual cognitive capacities, i.e., top-down abilities, have been shown to affect both the speech and listening effort outcomes in CI users (Pisoni, 2000; Amini et al.,

2023; Beckers et al., 2023). Working memory (WM) is a factor of most interest due to its association with speech performance in pediatric and adult CI users, HI and old NH listeners (Pisoni and Cleary, 2003; Akeroyd, 2008; Rudner et al., 2011; Besser et al., 2013; Füllgrabe and Rosen, 2016; Kestens et al., 2021; Dingemanse and Goedegebure, 2022). WM is a limited-capacity system with simultaneous storage and processing mechanisms whereby encoded information is analyzed and manipulated (Daneman and Carpenter, 1980). Studies using WM tasks incorporating either the processing speed or efficiency (for instance forward-and-backward digit span, reaction speed etc.) showed significant relation with CI users' speech recognition (Tao et al., 2014; Moberly et al., 2017a,b; Völter et al., 2021). Nonverbal reasoning tests have also been suggested as a method to predict speech recognition in CI users, because the ability to solve new tasks using limited information should, intuitively, relate to the ability to fuse degraded auditory inputs into a meaningful percept. Past studies have reported a significant correlation between CI users' sentence recognition and their non-verbal reasoning skills, although the underlying relation between nonverbal reasoning and processing efficiency remains unclear (Carpenter et al., 1990; Mattingly et al., 2018; Moberly et al., 2019; Beckers et al., 2023). Cognitive inhibition is the ability to suppress distracting information, and past studies have shown that this ability could be linked with CI word and sentence recognition (Moberly et al., 2018; Tamati et al., 2020; Völter et al., 2021; Tamati and Moberly, 2022). Considering the widely reported impact of individual cognitive abilities on CI outcomes, it is important to understand how CI users' individual cognitive capacities affect the relation between speech performance and listening effort, to choose the most optimal test levels for each CI user.

Therefore, the present study aims to, firstly, investigate the non-linear relationship between pupil response and speech performance in noisy conditions for CI users. Speech recognition and listening effort (measured by both concurrent pupillary response and subjective ratings) were tested in a wide range of SNR levels across the entire psychometric function of CI users, similarly to (Ohlenforst et al., 2018). We hypothesize that a similar inverse U-shape relation between speech performance and pupillary response would appear consistently in CI users, albeit with higher individual variability compared to NH and HI listeners.

The second aim of the present study is to identify sources and patterns in CI users' profiles that allow us to explain individual variabilities in the non-linear relationship mentioned above. The present study evaluates specifically the effect of neurocognitive factors (i.e., top-down) and auditory sensitivity (i.e., bottom-up), as measured by visual word-color (Stroop, 1935) (i.e., cognitive inhibition), N-back (Gevins and Cutillo, 1993) (i.e., working memory), Progressive Matrices (Harris et al., 2020) (i.e., non-verbal intelligence), and Spectral Modulation Ripple Test (SMRT) (Aronoff and Landsberger, 2013) (i.e., spectral-temporal sensitivity). We hypothesize that individual differences in CI users' performance and top-down/bottom-up profiles predict the shape of the psychometric relation between SNR and pupillary response.

To sum up, results of the present study will enrich our knowledge on the variable outcomes of speech perception and listening effort in CI users. It will help future research and clinics to interpret better the pupillometry outcomes when applying this method to quantify listening effort in CI users and assess innovative CI strategies.

2 Materials and methods

This section describes: (1) the characteristics of the recruited patients and (2) the experimental design and measures collected, which include: (a) clinical fitting and audiometry testing, (b) a hearing in noise test (HINT) as auditory stimuli to elicit pupil responses, (c) a setup for pupillometry recording to objectively quantify listening effort; (d) self-reported measures of subjective effort and cognitive load; (e) a spectro-temporal resolution modulation test; (f) cognitive tests to evaluate working memory, inhibitory control and non-verbal intelligence, and (g) quality of life. Finally, (3) data pre-processing and statistical methods are described.

2.1 Participants

Seventeen subjects (10 females and 7 males) with a mean age of 53.8 ± 12.2 years (range 36–69 years) volunteered to participate in the study from January to December 2022. All participants had been diagnosed with post-lingual severe-to-profound hearing loss and were unilateral Neuro Zti (Oticon Medical®) users with at least 12 months of experience with their CI (40.2 ± 11.1, range 12–59 months). Further demographic and clinical data are listed in Table 1. All the participants had normal or corrected-to-normal vision and no history of neurological or psychiatric diseases. The study was approved by the Ethics Committee of the Hospital Universitario Virgen Macarena in Seville, Spain (PEIBA#19/2019). All participants have written informed consent.

2.2 Experimental procedure

2.2.1 Clinical fitting and testing

Electrode impedances and maximum comfortable C levels were checked prior to the experimental speech test and pupillometry recording using the CI-link interface and the Genie Medical software (Oticon Medical®). All the participants wore a Neuro 2 sound processor with their most-frequently used fitting map. The noise reduction setting was deactivated to control behaviors of the automatic noise cancelation algorithm in laboratory conditions (Guevara et al., 2016). The compression system was set to the static version to avoid variable compression performance across the wide range of SNR conditions (Bozorg-Grayeli et al., 2016). Subsequently, participants were subjected to free-field aided audiometry from 250 Hz to 8,000 Hz with narrow-band stimuli (for detailed analysis on CI users' audiometry tested, see Supplementary Material 1). Finally, their word recognition was evaluated in quiet through a list of 25 Spanish disyllabic words (de Cárdenas and Marrero, 1994).

2.2.2 Stimuli

Auditory stimuli were Castilian Spanish sentences from the Hearing in Noise Sentence Test (HINT), embedded in speech-shaped noise matching the long-term spectrum of the sentence materials. Although the present study aims to target the complete psychometric range as in Ohlenforst et al. (2017, 2018), only a subset of SNRs was chosen because typical daily listening scenarios have SNRs ranging from 5 to 15 dB SNR (Smeds et al., 2015) and CI users show different ranges and slopes of their speech psychometric functions (Bergeron

TABLE 1 Demographic information of CI participants.

Participant	Age	Sex	Ear	Implant type	Array	Time of use (months)	Etiology
1	65	M	RIGHT	Neuro Zti	EVO	32	Unknown
2	60	F	LEFT	Neuro Zti	EVO	35	Sudden HL
3	44	F	RIGHT	Neuro Zti	EVO	42	Unknown
4	65	F	LEFT	Neuro Zti	EVO	45	Ototoxic
5	64	M	LEFT	Neuro Zti	EVO	32	Unknown
6	69	F	LEFT	Neuro Zti	EVO	41	Unknown
7	45	M	LEFT	Neuro Zti	EVO	32	Possible hepatitis
8	52	F	RIGHT	Neuro Zti	EVO	37	Hepatitis disease
9	39	F	LEFT	Neuro Zti	EVO	35	Unknown
10	45	M	LEFT	Neuro Zti	EVO	50	Unknown
11	61	M	LEFT	Neuro Zti	EVO	51	Unknown
12	53	M	RIGHT	Neuro Zti	EVO	42	Head trauma
13	59	F	RIGHT	Neuro Zti	EVO	39	Unknown
14	27	F	RIGHT	Neuro Zti	EVO	12	Unknown
15	63	M	LEFT	Neuro Zti	EVO	58	Unknown
16	48	F	RIGHT	Neuro Zti	EVO	59	Meniere disease
17	36	F	RIGHT	Neuro Zti	EVO	42	Unknown

and Hotton, 2016; Cordeiro et al., 2021). CI users also experience near-floor performance in negative SNRs, therefore not useful to inform on the psychometric shape of intelligibility and effort. Therefore, the present study tested 6 steps from 0 to +20 dB SNR in increment of 4 dB SNR: 0 dB, 4 dB, 8 dB, 12 dB, 16 dB, 20 dB SNR.

All stimuli were presented through a frontal loudspeaker located at 0° azimuth from the participant, who was sat on a chair located approximately 120 cm ±3 cm from the speaker in a soundproof room. Noise levels for all SNR conditions were fixed at 65 dB SPL, and speech levels varied accordingly. Note that due to the wide range of SNR tested in the current study, prior to the experiment, a quick test was administered to ensure that speech level at 0 dB SNR was audible, and overall loudness at 20 dB SNR was loud but tolerable. CI users were presented with an initial training list of 20 sentences at a maximum SNR of +20 dB SNR. Subsequently, different lists of 20 sentences were presented, twice per different SNR conditions. The presentation of the SNR conditions was randomized. Overall, 240 sentences (40 sentences per 6 SNR conditions) were tested per participant.

2.2.3 Concurrent pupillometry

Pupil diameter was recorded using an infrared eye tracking system PupilLabs Core[1] (Kassner et al., 2014). Pupil is an open-source platform for mobile eye-tracking and gaze-based measurements. The experimenter monitored and ensured the quality and stability of the real-time pupil responses recorded before and during the session, as well as other settings such as the camera exposure and intensity parameters available in the eye-tracker Pupil GUI software. Participants were instructed to fixate their eyes on a fixation cross at the TV screen and were asked not to move or blink brusquely as much

as possible. Room luminance was controlled by adjusting the room ambience light and screen brightness to reach an average value around 55 ± 5.1 lux, using a luxmeter positioned at the eye level of the participant facing the screen. The luminance level was then kept stable throughout the recording to ensure no impact on pupil traces from non-task-evoked pupillary response (Zhang et al., 2022). The eye-tracker device was connected to a laptop and the pupil data was collected through customized Python scripts developed within the OMEXP research platform (Sulas et al., 2022).

In each trial, each participant was instructed to fixate on a light gray fixation cross (+) on the dark gray screen. The background noise started 2 s before the sentence's onset, serving as a baseline measurement. The background noise then ended 2 s after the offset of the sentence, allowing for task-evoked pupillary response to peak (Winn et al., 2018). After the offset of background noise, the fixation cross changed to a cross (X), indicating the participant to repeat the sentence aloud. The experimenter noted down the correctly repeated keywords in the sentence then moved on to the next trial. Pupil recordings were synchronized with HINT sentences at interval points corresponding to the start of each trial (noise onset), sentence onset and post-noise offset using the Lab Streaming Layer (LSL) protocol (Kothe et al., 2012). To avoid fatigue in some patients, the experiment was run in two different experimental sessions on two separate days (at around the same time of the day) if needed. Note that past studies have shown that task performance and pupil responses can be situationally dependent and can vary by the day (Veneman et al., 2013; Winn et al., 2018). However, as shown later in the analysis section, the current experiment corrected pupillary response by the trial baseline which could alleviate the day-to-day variation in absolute pupil diameter. Therefore, to control for the confounding effect of fatigue on task-evoked pupil responses, the test organization was prioritized to avoid fatigue (McGarrigle et al., 2017; Zekveld et al., 2018).

---

1 https://docs.pupil-labs.com/

2.2.4 Subjective measures of mental workload and effort

After each SNR condition (20 sentences), subjects were asked to report their subjective mental workload using a computer version of the well-known NASA Task Load Index (TLX) (Hart and Staveland, 1988). This yields a multimodal overall workload score calculated from a weighted average of subscale ratings including six categories: physical demand, temporal demand, mental demand, self-performance, effort, and frustration. The test comprised two phases: 1) from two categories, the patient selected the one that best reflected the sensation experienced during the task. All combinations of two categories were tested; 2) for each category, a visual analog scale ranging from "low" to "high" or from "good" to "poor" (in the case of performance) was presented to the user. Scores were calculated following the standard procedure.

2.2.5 Spectro-temporal resolution

To quantify spectral resolution, the SMRT (Aronoff and Landsberger, 2013) was administered with the same setting conditions, at 65 dB SPL from a frontal loudspeaker located at 0° azimuth from the participant. In each trial, participants were presented with three stimuli (2 references; 1 target) and were asked to press a button to indicate which of the three stimuli sounded different. The references were fixed at 20 ripples per octave (RPO). The target initially had a ripple density of 0.5 RPO. The ripple density of the target varied adaptatively depending on user responses using a one-up/one-down procedure and a step size of 0.2 RPO. The test stopped after ten reversals, and the last six reversals were averaged to give an overall SMRT threshold (in RPO) for each participant.

2.2.6 Neurocognitive functioning

Working memory was assessed by the visual N-back test with N = 2 which involves a sequence of alphabets presented in a computer screen. In the 2-back test, subjects were asked to press a key on a keyboard each time an alphabet shown on the screen was identical to that shown two alphabets before. Each run comprised 22 alphabets and each participant took two runs. Discriminability d-prime was calculated using R package psycho (Makowski, 2018), as an index of working memory efficiency (Haatveit et al., 2010).

Inhibitory control was assessed using Stroop color-word visual test based on the original version (Stroop, 1935). During each trial, a color word was presented on a computer screen, presented in a specific font color, and the participant was asked to press a key on a keyboard corresponding to the font color and not the name of the color word. The test contains three distinct types of trials: congruent trials (color font and color word match), incongruent trials (color font and color word do not match), and control trials (a rectangle of a given color is shown). Each type was blocked and comprised of 12 unique trials. Each block was repeated 3 times, and the block sequence was randomized. Stroop color-word interference scores were calculated by subtracting the mean correct of the incongruent trials by the congruent trials. The average reaction time in control trials was also calculated for later analysis.

Lastly, nonverbal intelligence was measured using a computer version of Progressive Matrices, which shows a series of progressive 3 × 3 visual grid patterns that become increasingly difficult in time (Matzen et al., 2010; Harris et al., 2020). Participants were asked to complete the pattern by selecting the last unfilled grid from several options shown. Participants were scored based on how many items they correctly completed.

This set of neurocognitive tests intended to capture the differences in CI users' abilities. N-back task assessed CI users' abilities to store and process the activated phonological and lexical competitors from the ambiguous inputs; Stroop color-word test assessed their abilities to reject incorrect candidates; and Progressive Matrices task assessed their abilities to fuse ambiguous information into a complete percept.

The spectro-temporal and neurocognitive tests were also performed using the OMEXP computer platform based on custom-made and validated Python scripts (Sulas et al., 2022).

2.2.7 Quality of life

Quality of life (QoL) was evaluated through the Nijmegen Cochlear Implant Questionnaire (NCIQ) (Hinderink et al., 2000). NCIQ is a well-known and commonly used test to evaluate auditory QoL in CI users. NCIQ comprises three general domains (physical, psychological, and social), which can be divided into six different subdomains: the physical domain is formed by three subdomains: basic sound perception (NCIQ1), advanced sound perception (NCIQ2) and speech production (NCIQ3). The psychological domain is formed by the subdomain self-esteem (NCIQ4), and the social domain comprises two subdomains; activity (NCIQ) and social interactions (NCIQ6). Each of these subdomains comprises ten questions or statements answered through a five-point response scale ranging from 'never' to 'always' or from 'no' to 'good'. If the question does not apply to a patient, a sixth answer 'not applicable' is also available. Final overall scores range from 0 (very poor) to 100 (optimal) for each subdomain and globally.

2.3 Data pre-processing and analysis

2.3.1 Pupil data pre-processing

Pupil traces collected from the left-eye camera were pre-processed to detect eye blinks and abnormalities. Any sample point with a value below or above 3 Median Absolute Deviation (MAD) of the median pupil diameter of the recording was counted as blinks or outliers and interpolated using cubic interpolation (Leys et al., 2013). Additionally, sample points 35 ms before and 100 ms after the abnormality were also discarded to avoid blink onset and offset effects. The missing samples were cubically interpolated based on the preserved samples of the trace and smoothed with a running average filter with 100-ms duration. Trials with over 40% of the sample points interpolated were excluded from further analysis. A previous study also supported that 45% blink exclusion criterion would not affect group pupil results compared to more stringent criteria (Burg et al., 2021). On this basis, 12% of all trials were excluded, and no participant was excluded because no participant had over 30% trials excluded.

The baseline pupil diameter in each trial was calculated as mean pupil diameter averaged over 2 s before the onset of the sentence. The pupil diameter measured from the sentence onset to the repeat prompt was subtracted from that baseline level and then normalized by the baseline diameter to obtain the relative pupil diameter changes elicited by the task, according to the formula below (Wagner et al., 2019):

$$\frac{observation - baseline}{baseline} \times 100 \qquad (1)$$

Processed traces were then aggregated per participant and per SNR condition, aligned by the onset of the sentences. The search for the maximum pupil dilation was restricted to the time window starting from sentence onset and ending at the verbal response prompt (thus excluding any possible confounding effect with the pupil arousal induced when verbally repeating the sentence), and the maximum dilations were recorded as PPD for later analyses. Note that all sentences were left unprocessed in duration to avoid unnatural acoustic manipulation (compression or stretching of original sentences). This procedure was common in listening effort studies due to varied length of standardized sentences and the variability in sentence duration could be well controlled by consistent trace alignment (either by sentence onset or offset). All pupil data pre-processing was conducted on Matlab 2016b using custom-made scripts.

2.3.2 Statistical analysis

To examine the first hypothesis that a consistent psychometric relation similar to that of NH and HI listeners exists for CI users, a mixed-effect linear regression model was built, using PPD as dependent variable, SNR condition as the fixed effect factor, and participant as random effect factor. The model was constructed using the lme4 package (Bates et al., 2015) in R and figures were produced using the ggplot2 package (Wickham, 2016). Fixed and random effect factors entered the model, and remained in the model only if they significantly improved the model fitting, using Chi-squared tests based on changes in deviance ($p < 0.05$). Differences among levels of SNR condition were examined using post-hoc Wald test, and p values were estimated using the z distribution in the test as an approximation for the t distribution (Mirman, 2017). To perform sanity checks on whether behavioral performance and subjective rating in speech-in-noise tasks were consistent with past CI results, two mixed-effect logistic regression models were built, using keywords proportion correct and NASA-TLX scores as dependent variables respectively, SNR condition as the fixed effect factor and participant as random effect factor.

Considering that we hypothesized an inverted-U shape for the psychometric relation between task difficulty and pupillary response, a mixed-effect linear regression model with linear and quadratic polynomial terms was constructed. The model used PPD as dependent variable, linear and quadratic term of SNR condition as fixed effect factors and participant as random effect factor, to examine whether linear (i.e., rising or falling) and quadratic trends (i.e., U-shape or inverse U-shape) existed from 0 dB SNR to 20 dB SNR. Model selection and *post-hoc* tests were performed in the same way as above.

To examine the second hypothesis that individual variability could predict the shape of the psychometric relation, a series of correlation tests were performed. For parameters relating to the shape of the psychometric relation, average PPD and quadratic coefficients of the polynomial fitting of PPD from SNR0 to SNR20 were calculated for each participant. For measured individual differences in CI performance and top-down/bottom-up factors, age, CI duration, clinical word recognition in quiet, auditory thresholds measured in free field (T-level), QoL, sentence keyword corrects in noise, NASA-TLX rating in noise, Progressive Matrices correct, Stroop inference scores in correct, Stroop control reaction time, N-back d-prime and SMRT threshold were considered. Pairwise correlation

tests were performed to examine the strength of correlation between individual differences and psychometric shape parameters. Pearson's r was used when two factors under comparison were continuous, and Spearman's rho was used when either factor was non-continuous (for instance, values bounded between 0 and 100%) or non-normal bivariate distribution. Bonferroni correction on p-values was applied to correct for multiple testing.

3 Results

3.1 Sentence recognition and subjective effort rating

A significant effect of SNR condition on sentence keyword recognition score was found ($\chi^2 = 118.3$, df = 5, $p < 0.001$). Post-hoc z-tests showed a consistent trend that higher SNRs were related to better recognition scores. Statistically significant comparisons are: SNR12 > SNR4 ($\beta = 4.33$, SE = 1.12, $p < 0.001$), SNR16 > SNR4 ($\beta = 6.41$, SE = 1.53, $p < 0.001$), SNR20 > SNR4 ($\beta = 6.96$, SE = 1.61, $p < 0.001$), SNR12 > SNR8 ($\beta = 2.81$, SE = 1.01, $p = 0.005$), SNR16 > SNR8 ($\beta = 4.89$, SE = 1.4, $p < 0.001$), SNR20 > SNR8 ($\beta = 5.43$, SE = 1.49, $p < 0.001$), SNR20 > SNR12 ($\beta = 2.63$, SE = 1.33, $p = 0.03$).

A significant effect of SNR condition on NASA-TLX rating was found ($\chi^2 = 21.62$, df = 5, $p < 0.001$). Post-hoc z-tests showed a consistent trend that higher SNRs were related to easier ratings. Statistically significant comparisons are: SNR12 < SNR0 ($\beta = -0.81$, SE = 0.26, $p = 0.002$), SNR16 < SNR0 ($\beta = -0.81$, SE = 0.26, $p = 0.002$), SNR20 < SNR0 ($\beta = -0.86$, SE = 0.26, $p < 0.001$), SNR12 < SNR4 ($\beta = -0.53$, SE = 0.26, $p = 0.04$), SNR16 < SNR4 ($\beta = -0.53$, SE = 0.25, $p = 0.04$), SNR20 < SNR4 ($\beta = -0.58$, SE = 0.25, $p = 0.02$), SNR12 < SNR8 ($\beta = -0.59$, SE = 0.25, $p = 0.02$), SNR16 < SNR8 ($\beta = -0.58$, SE = 0.25, $p = 0.02$), SNR20 < SNR8 ($\beta = -0.64$, SE = 0.25, $p = 0.01$). Figure 1 shows the results of sentence keyword recognition (Figure 1A) and NASA-TLX score (Figure 1B).
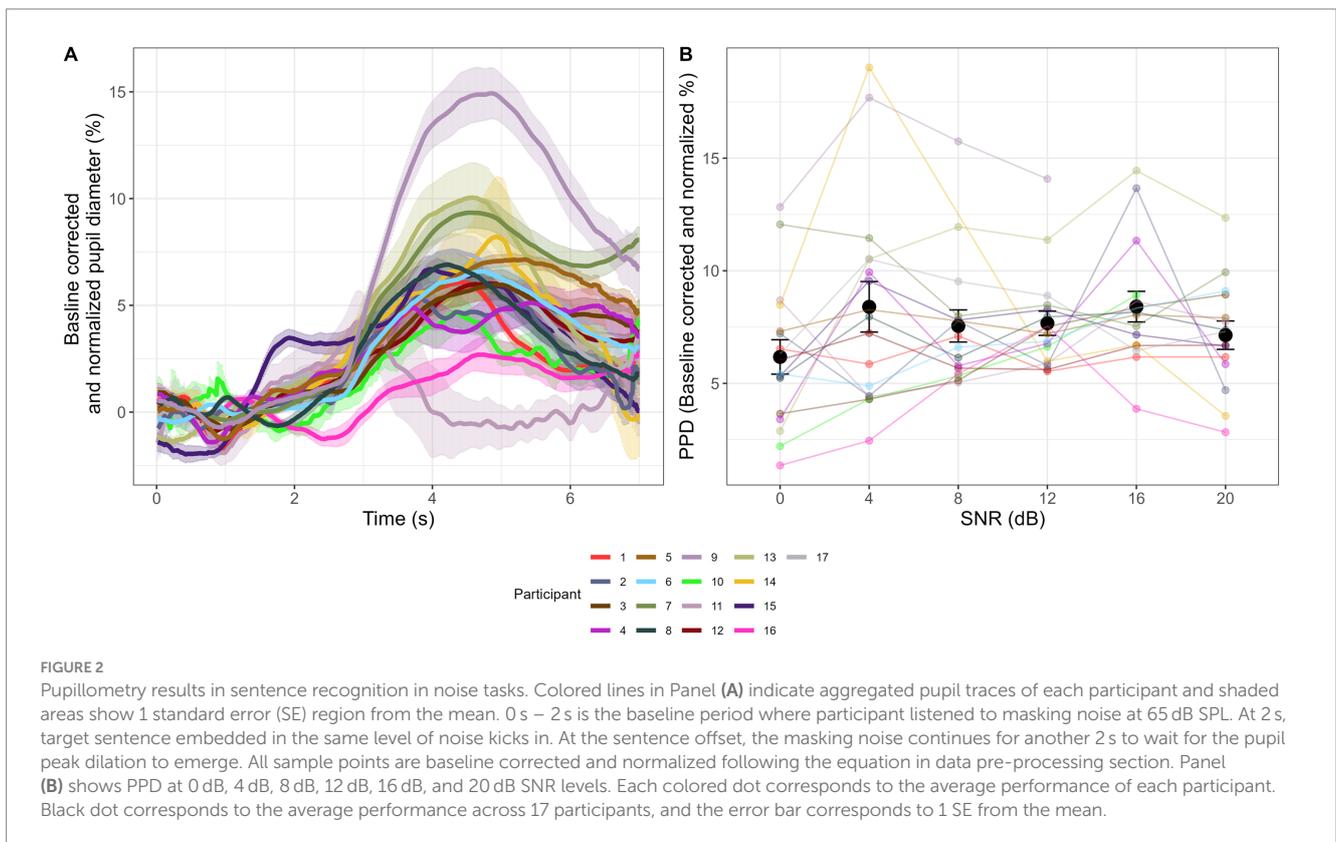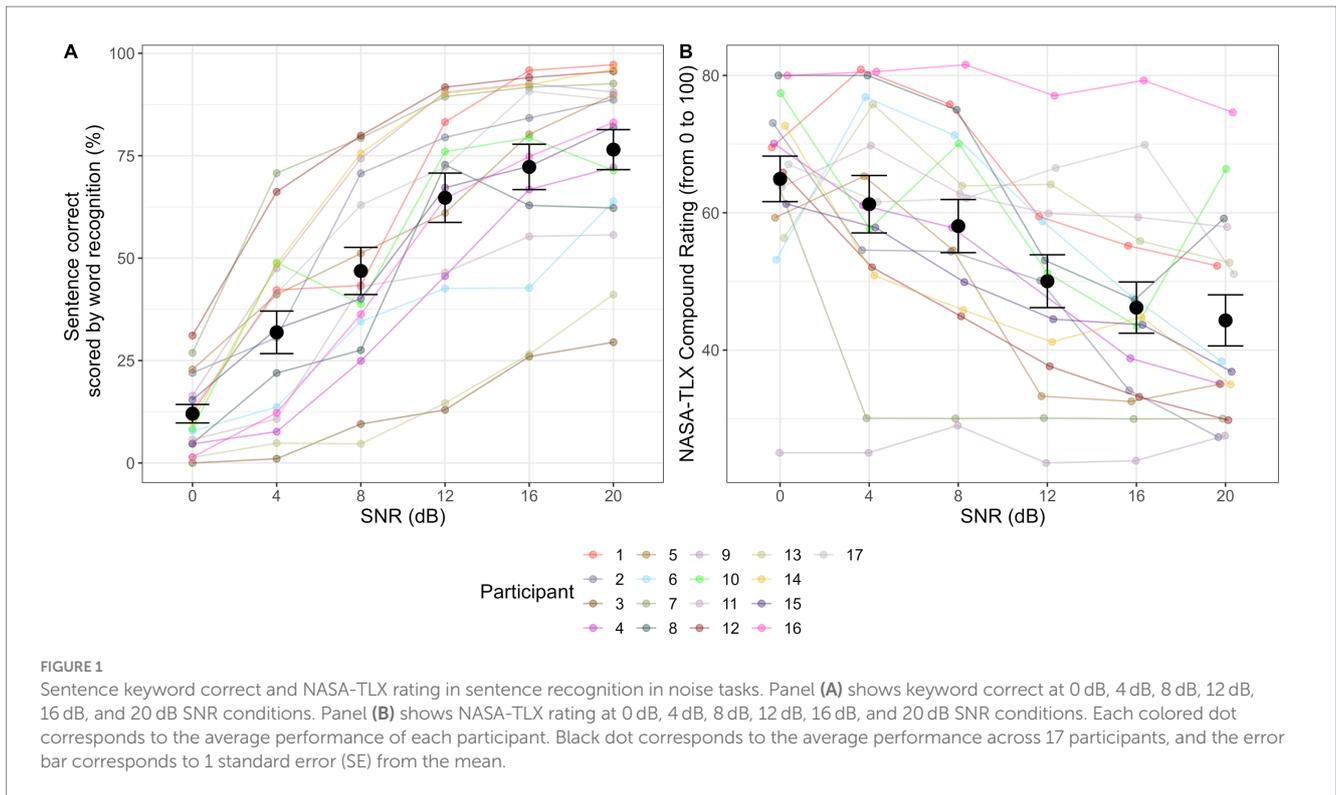
3.2 PPD and psychometric relation

Figure 2 shows the aggregated pupil traces (Figure 2A) and PPD across SNR levels (Figure 2B).

A significant effect of SNR condition on PPD was found ($\chi^2 = 13.23$, df = 5, p = 0.02). Post-hoc z-tests showed significant comparisons as: SNR4 > SNR0 ($\beta = 0.022$, SE = 0.008, $p = 0.007$), SNR16 > SNR0 ($\beta = 0.028$, SE = 0.008, $p < 0.001$).

Polynomial shape analysis showed no significant linear trend in PPD from SNR0 to SNR20 ($\chi^2 = 0.89$, df = 1, $p = 0.35$, $\beta = 0.04$, SE = 0.04, $p = 0.34$), and no significant quadratic trend ($\chi^2 = 3.82$, df = 1, $p = 0.051$, $\beta = -0.06$, SE = 0.03, $p = 0.056$), suggesting that at a group level there was no consistent increasing/falling or U-shape/inverse U-shape.

3.3 Individual variances in predicting the psychometric curve

Table 2 shows results of correlation tests between individual factors and performances in speech in noise recognition task. Figure 3 shows plots of significant correlations, after Bonferroni correction. For

FIGURE 1
Sentence keyword correct and NASA-TLX rating in sentence recognition in noise tasks. Panel **(A)** shows keyword correct at 0 dB, 4 dB, 8 dB, 12 dB, 16 dB, and 20 dB SNR conditions. Panel **(B)** shows NASA-TLX rating at 0 dB, 4 dB, 8 dB, 12 dB, 16 dB, and 20 dB SNR conditions. Each colored dot corresponds to the average performance of each participant. Black dot corresponds to the average performance across 17 participants, and the error bar corresponds to 1 standard error (SE) from the mean.

FIGURE 2
Pupillometry results in sentence recognition in noise tasks. Colored lines in Panel **(A)** indicate aggregated pupil traces of each participant and shaded areas show 1 standard error (SE) region from the mean. 0 s – 2 s is the baseline period where participant listened to masking noise at 65 dB SPL. At 2 s, target sentence embedded in the same level of noise kicks in. At the sentence offset, the masking noise continues for another 2 s to wait for the pupil peak dilation to emerge. All sample points are baseline corrected and normalized following the equation in data pre-processing section. Panel **(B)** shows PPD at 0 dB, 4 dB, 8 dB, 12 dB, 16 dB, and 20 dB SNR levels. Each colored dot corresponds to the average performance of each participant. Black dot corresponds to the average performance across 17 participants, and the error bar corresponds to 1 SE from the mean.

complete distribution and correlation visualizations for all individual factors, see Supplementary Material 2.

To further interpret the consistent change and significant correlation between clinical word recognition and the progression of the quadratic term of PPD from SNR0 to SNR20 (Figure 3C, rho = −0.76, p = 0.0004), 17 participants were split into two groups from the median based on their clinical word recognition performance. Same data used to plot Figure 2B were used to plot the

TABLE 2 Table with correlation test results.

	Age	CI dur	Word Rec	T-Level	QoL	NASA-TLX	Sentence correct in noise	Matrices correct	Stroop correct interference	Stroop control reaction time	Nback dprime	SMRT threshold
PPD	t = −1.17, r = −0.29, p = 0.26	t = −1.22, r = −0.3, p = 0.24	S = 530.3, rho = 0.35, p = 0.17	t = 0.94, r = 0.24, p = 0.36	**t = 3.99, r = 0.72, p = 0.001**	S = 1,050, rho = −0.54, p = 0.03	S = 592, rho = 0.27, p = 0.29	**S = 1378.7, rho = −0.69, p = 0.002**	S = 520.18, rho = 0.36, p = 0.15	t = 1.22, r = 0.30, p = 0.24	t = −0.37, r = −0.95, p = 0.72	t = 1.72, r = 0.41, p = 0.11
Quadratic term of PPD psychometric curve	t = 2.04, r = 0.47, p = 0.06	t = 1.81, r = 0.42, p = 0.09	**S = 1433.5, rho = −0.76, p = 0.0004**	t = −0.53, r = −0.13, p = 0.61	t = −2.10, r = −0.48, p = 0.05	S = 308, rho = 0.55, p = 0.03	**S = 1,442, rho = −0.77, p = 0.0003**	S = 502.58, rho = 0.38, p = 0.13	S = 858.26, rho = −0.05, p = 0.84	t = 0.02, r = 0.006, p = 0.98	t = 0.47, r = 0.12, p = 0.64	t = −3.09, r = −0.62, p = 0.007

Each column contains shape parameters relating to the psychometric relation between task difficulty (SNR) and pupillary response (PPD). Each row contains individual performance and top-down/bottom-up variabilities. Results in bold are correlation tests that remain significant after Bonferroni correction ($p < 0.002$).
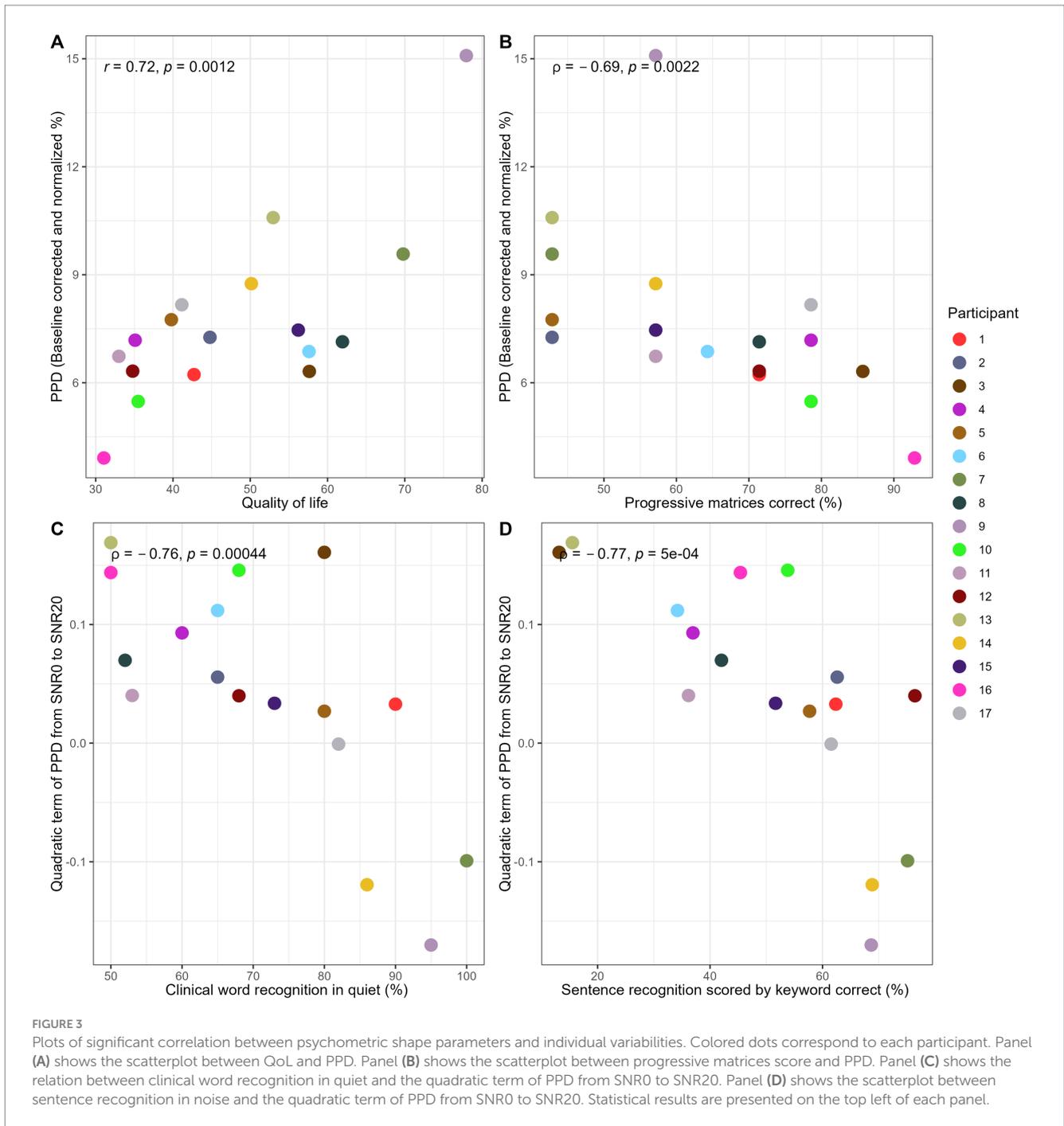
results of two sub-groups in Figure 4. Same analysis methods used for investigating the first hypothesis (i.e., whether a consistent psychometric relation exist) was performed on two groups, to validate visual observations in Figure 4. Accordingly, this additional fixed effect factor termed wordgroup with two levels (1st, 2nd quantile and 3rd, 4th quantile of word recognition) was entered into the mixed effect models. Same model and post-hoc testing methods were performed to examine the main effect of wordgroup and its interaction with SNR condition on PPD and polynomial shapes. Note that this sub-group analysis intends to visualize and support the interpretation of a significant across-participant correlation that is already reported within the scope of our second hypothesis (i.e., significant correlation between PPD psychometric curve quadratic term and word recognition (rho = −0.76, p = 0.0004)). Therefore, this extra investigation does not impair the statistical integrity of the main analysis by increasing the chances of false positives. Splitting from the median of the clinical word recognition would not bias our interpretation, considering that there was a consistent across-participant correlation. (For sanity check on this point, see Supplementary Material 3 where CI participants are split into 1st and 4th quantile. With fewer participants and more different word recognition performance in each sub-group, results remain similar to using median split.) Independent sample t-tests verified that the two sub-groups did not differ in other profiles such as age ($t = −1.53$, $df = 10.51$, $p = 0.16$), mean T-level ($t = 1.45$, $df = 14.20$, $p = 0.17$), CI duration ($t = −1.31$, $df = 12.27$, $p = 0.21$), QoL ($t = 1.82$, $df = 13.84$, $p = 0.09$), SMRT threshold ($t = 1.81$, $df = 14.73$, p = 0.09), Nback d-prime ($t = 0.73$, $df = 13.58$, $p = 0.48$), Stroop interference score ($t = 0.8$, $df = 9.23$, $p = 0.44$), nor speech in noise task performance ($t = 1.40$, $df = 14$, $p = 0.18$).

There was no significant main effect of wordgroup on PPD ($\chi^2 = 2.69$, $df = 1$, $p = 0.10$), suggesting that the two subgroups did not differ in average PPD. There was a significant interaction between SNR and wordgroup on PPD ($\chi^2 = 13.72$, $df = 5$, $p = 0.02$). In the group where participants were in the 1st and 2nd quantile of clinical word recognition, SNR4 > SNR0 ($\beta = 0.03$, $SE = 0.01$, $p = 0.01$), SNR16 > SNR4 ($\beta = −0.03$, $SE = 0.01$, $p = 0.03$), SNR20 > SNR4 ($\beta = −0.02$, $SE = 0.01$, $p = 0.04$). In the group where participants were in the 3rd and 4th quantile of clinical word recognition, SNR12 > SNR0 ($\beta = 0.02$, $SE = 0.01$, $p = 0.03$), SNR16 > SNR0 ($\beta = 0.05$, $SE = 0.01$, $p < 0.001$), SNR20 > SNR0 ($\beta = 0.02$, $SE = 0.01$, $p = 0.04$), SNR16 > SNR4 ($\beta = 0.03$, $SE = 0.01$, $p = 0.005$), SNR16 > SNR8 ($\beta = 0.03$, $SE = 0.01$, $p = 0.008$), SNR16 > SNR12 ($\beta = 0.02$, $SE = 0.01$, $p = 0.02$), SNR20 < SNR16 ($\beta = −0.025$, $SE = 0.01$, $p = 0.01$). All the post-hoc tests were consistent with the pattern in Figure 4.

There was a significant interaction between wordgroup and the linear shape ($\chi^2 = 6.03$, $df = 1$, $p = 0.01$), suggesting that high word recognition group had more falling trend than low word group progressing from SNR0 to SNR20 ($\beta = 0.18$, $df = 11.67$, $p = 0.03$).

4 Discussion

In this study, with CI users we revisited the well-established psychometric relation between speech recognition and pupillary response reported in NH and HI listeners (Ohlenforst et al., 2017, 2018; Wendt et al., 2017). By replicating the experiment on a group of post-lingually deaf CI users, we constructed a similar psychometric

FIGURE 3
Plots of significant correlation between psychometric shape parameters and individual variabilities. Colored dots correspond to each participant. Panel **(A)** shows the scatterplot between QoL and PPD. Panel **(B)** shows the scatterplot between progressive matrices score and PPD. Panel **(C)** shows the relation between clinical word recognition in quiet and the quadratic term of PPD from SNR0 to SNR20. Panel **(D)** shows the scatterplot between sentence recognition in noise and the quadratic term of PPD from SNR0 to SNR20. Statistical results are presented on the top left of each panel.

function across a wide range of SNR and intelligibility levels. We also measured individual variabilities in audiological status, auditory sensitivity, and cognitive profile, to examine their strength in predicting individual psychometric curves.

4.1 PPD relates to listening effort (in the right region of the PPD psychometric curve)

At a group level, CI results showed no consistent inverse U-shape as in NH and HI listeners. To aid direct comparison with previous studies, Figure 5A joins the information in Figure 1A and Figure 2B to re-plot the same double psychometric function figure as in previous

studies (Ohlenforst et al., 2017, 2018; Wendt et al., 2017). Figure 5A shows no consistent increasing/decreasing trend or U-shape/inverse U-shape in PPD, from 0 dB to 20 dB SNR, which covers a wide range of difficulty and intelligibility. Note that failure to find the same relation is not due to bias in participants selection or errors in experimental design. CI users in this study show comparable audiological profile, speech-in-noise, and clinical word recognition performance as the general CI population (Holden et al., 2013; Goudey et al., 2021). Pupil traces and obtained PPD are also comparable with past studies (Winn et al., 2015; Wagner et al., 2019; Russo et al., 2020). Furthermore, consistent subjective effort ratings indicate that participants experience distinctive levels of effort across different conditions. Therefore, our experimental manipulation has
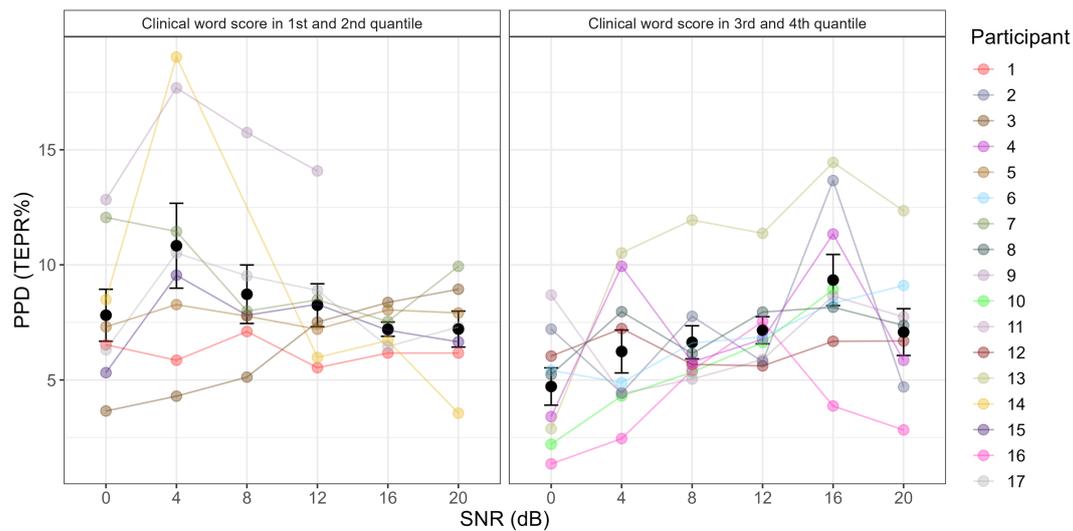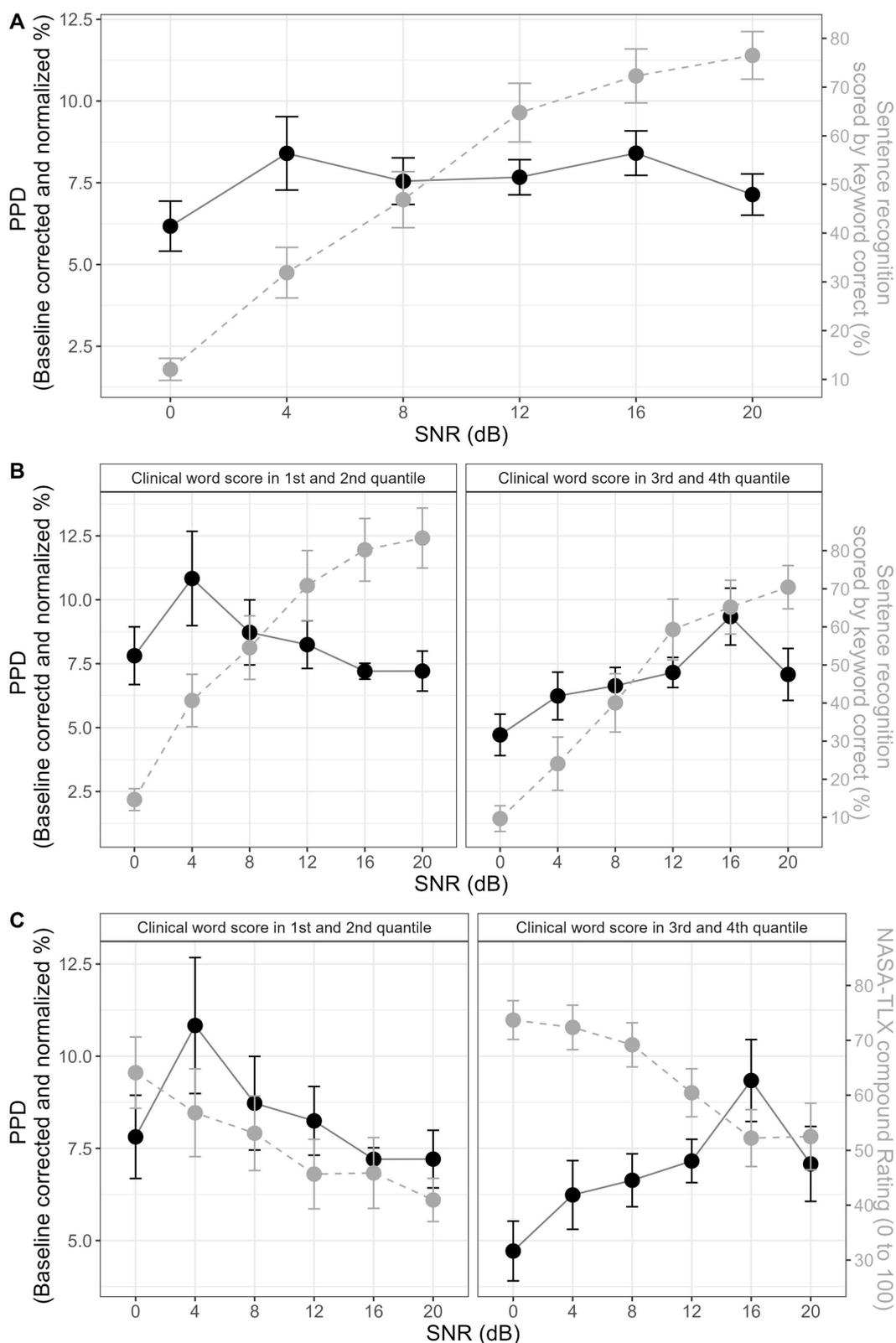
FIGURE 4
PPD results at different SNR levels, with participants split into two groups. Left panel shows results for participants whose clinical word scores are in the 1st and 2nd quantile, and 4 dB SNR condition was significantly higher than 0 dB, 16 dB, and 20 dB SNR. The right panel shows results for participants whose clinical word scores are in the 3rd and 4th quantile, and 16 dB SNR condition was significantly higher than 0 dB, 4 dB, 8 dB, 12 dB, and 20 dB SNR.

succeeded in varying intelligibility and perceived listening effort, yet the effect is not shown in PPD.

Anecdotally, there have been reports from clinical and research sites that pupillometry has not been reliably reflecting changes in experimental manipulations for CI users, casting doubts on the suitability of pupillometry as a tool for CI applications. And our results seem to agree with the anecdotal reports from the field. This is worrying, because if pupillary response does not show any consistent variation pattern across various levels of perceived listening effort in CI users, then pupillometry cannot be used as a tool to reliably assess listening effort and inform research and clinical practice.

However, when the great individual variability in clinical word performance is considered, results show a significant correlation between psychometric curve and clinical word performance (rho = −0.76, p = 0.0004). Similarly, to aid interpretation and direct comparison with previous studies, Figure 5B joins the information in Figures 1A, 3 to re-plot the double psychometric curves. Again, as emphasized in the analysis, although Figures 4, 5B arbitrarily split CI participants into two groups based on the median clinical word recognition, the significant inter-participant correlation indicates that the pattern shown in the two figures is consistent from the lowest to the highest CI performers. A significant correlation between sentence recognition in noise and psychometric curve further supports this pattern (rho = −0.77, p = 0.0003). This suggests that pupillometry data in CI users are not entirely chaotic but driven by individual differences in speech intelligibility. As shown in Figure 5B, CI users display a similar inverse U-shape in PPD psychometric curve as NH and HI listeners but differ in the location of the "tipping point" of the curves. In the lower quantile group (i.e., low word recognition), the "tipping point" occurs at 16 dB SNR or around 40% intelligibility, and in the higher quantile group (i.e., high word recognition), the "tipping point" occurs at 4 dB SNR or around 70% intelligibility. Not surprisingly, when combining the two groups into one figure (Figure 2B), the mean psychometric curve appears flat.

Furthermore, speech intelligibility may not be the only difference between NH/HI listeners and CI users. It is no news that the same SNR level does not yield same speech recognition scores among NH, HI, and CI listeners, so comparing directly 'tipping points' in terms of their corresponding SNR levels is not thorough enough. Regardless of the difference in SNR levels, NH and HI listeners typically showed the 'tipping point' on their pupillometry psychometric curves at around 40–50% speech intelligibility (Ohlenforst et al., 2017; Wendt et al., 2017; Dingemanse and Goedegebure, 2022). This suggests that even though NH and HI listeners have different speech recognition, they tend to disengage from the task at similar intelligibility level. But CI users show great difference in the intelligibility level where the 'tipping point' occurs (Figure 5B). It seems that depending on which quantile of clinical word recognition performance CI users land in, they disengage from the task at either high (~70%) or low intelligibility level (~40%) in speech-in-noise tasks. For instance, for CI users in lower quantile group (Figure 5B left panel), even at testing levels when sentence recognition is successful (say, >60%), PPD increases with increasing intelligibility or easier SNR. At the same speech intelligibility region, CI users in higher quantile group show the opposite trend, where PPD decreases with increasing intelligibility or easier SNR. Could it be that CI users also differ in subjective decision (consciously or unconsciously) of when to disengage from a task? This mechanism could explain the difference in the consistency of the "tipping point" observed in NH/HI listeners and in our CI users. For instance, even though relatively low-performing CI users can perform successful speech recognition, the cost of such an effort is deemed too huge to be rewarding or motivating, hence translating into an "inversion" in physiological markers and negative listening effort that could further evolve to listening fatigue. In our current study, NASA-TLX is a subjective measure of effort that does not include the dimension of motivation and engagement. At a group level, NASA-TLX also related to pupillometry psychometric curve, but did not pass Bonferroni correction (rho = −0.54, p = 0.03 for correlation

FIGURE 5
Double psychometric function figures for PPD, sentence recognition in noise and NASA-TLX. In all panels, dot corresponds to the average performance, and the error bar corresponds to 1 standard error (SE) from the mean. Panel **(A)** combines dots and error bars in black solid line to indicate PPD psychometric function, and gray dashed line to indicate sentence recognition psychometric function. Panel **(B)** splits participants into two group, with the left side containing participants whose clinical word recognition falls in the 1st and 2nd quantile and the right-side 3rd and 4th quantile. Gray dashed lines in panel **(C)** connect dots and error bars to indicate NASA-TLX psychometric functions.

with the average PPD; rho = 0.55, p = 0.03 for correlation with the quadratic term). A similar combination of Figures 1B, 3B shows that, interestingly, both 'tipping points' in the two groups occur at around 50% NASA-TLX score, a similar value at which NH/HI listeners disengage but in terms of objective intelligibility (Figure 5C). According to the Framework for Understanding Effortful Listening (FUEL), the allocation of limited cognitive resources for the listening task is influenced by the listener's motivation (Pichora-Fuller et al., 2016). Effort investment will increase with task demands, as long as there is motivation to succeed at the task and the goal is deemed attainable (Brehm and Self, 1989). Maximal effort is the point when the listener is sufficiently motivated to succeed in communication, and listening is demanding but not perceived as wasted effort (Richter, 2013, 2016; Pichora-Fuller et al., 2016; Herrmann and Johnsrude, 2020). It is likely that the 'tipping point' observed in PPD psychometric function could be the physiological biomarker of such maximal effort, where further task demands will result in de-motivation and task being deemed as not worthwhile. Past pupillometry studies have shown pupil dilation is sensitive to different motivation levels and conditions, due to its connection with the locus coeruleus noradrenergic system that is responsible for regulating task engagement (Murphy et al., 2011; Carolan et al., 2022). And we further hypothesize that pupillary response can be sensitive enough to reflect the threshold where task engagement starts to decrease. Note that the interpretation of the 'tipping point' as point of disengagement or 'giving up' has not been validated in past studies, for instance using a subjective questionnaire on motivation or engagement. Our current study does not have the correct subjective questionnaire for quantifying motivation or engagement, nor enough power to examine this interpretation statistically, due to a lack of a measure of motivation or engagement and the fear of inflated false positive rates. Therefore, future studies need to incorporate measures of motivation or engagement to uncover the ecological meaning of the 'tipping point' and validate its sensitivity and reliability as a measure of disengagement in clinical populations.

4.2 Individual factors affect individual PPD psychometric function

Despite the big individual differences in pupil psychometric curves, individual differences measured in life status and neurocognitive factors explain certain amount of variability significantly (mean R^2 = 0.54, p < 0.002 after Bonferroni correction). Non-verbal intelligence measured by Progressive Matrices is negatively correlated with the average PPD, suggesting that CI participants with higher non-verbal intelligence showed lower PPD across all SNR conditions. The relation between non-verbal intelligence and speech outcomes in CI users have been examined in past studies, but it was unclear how non-verbal intelligence affect listening effort (Mattingly et al., 2018; Moberly et al., 2019; Amini et al., 2023; Beckers et al., 2023). Non-verbal intelligence is typically found to relate positively with word or sentence recognition, indicating that higher ability to induce abstract relation helps recognizing sentences that have degraded resolution due to CI signal processing (Carpenter et al., 1990). Similarly, better non-verbal intelligence should also help relieve listening effort by listeners having larger cognitive capacities at disposal. However, no significant correlation

was found between non-verbal intelligence and quadratic terms of PPD psychometric curve (unlike clinical word and sentence-in-noise recognition), suggesting that non-verbal intelligence is not associated with the psychometric function curvature. It is possible that individual cognitive capacities can modulate the relation between listening effort and performance for a given task, but not necessarily the structural relation with the point of de-motivation or dis-engagement when facing various levels of difficulties.

QoL measured by NCIQ related positively with the average PPD, suggesting that CI participants with higher quality of life ratings showed higher PPD across all SNR conditions. This result seems to be counter intuitive and inconsistent with past findings (especially when there is a positive correlation between QoL and word recognition performance, see Supplementary Material 2). For instance, past studies have reported that higher Speech, Spatial and Qualities (SSQ) related with smaller pupillary response in CI users at 10 dB SNR and quiet conditions (Russo et al., 2020); higher subjective daily fatigue related with bigger PPD (Wang et al., 2018). Arguably, NCIQ contains more domains than subjective hearing status and daily fatigue, and includes measures on physical (basic sound perception, advanced sound perception, speech production), psychosocial (self-esteem) and social (activity limitations, social interactions) functioning, hence could be affected by various cognitive, sensory, and demographic factors (Skidmore et al., 2020). It is still unexpected that the direction of correlation is the opposite. One possibility could be that the average PPD, as a psychometric function shape parameter in our current study, reflects more than explicit cognitive resources allocated to the task, but also the multidimensionality of listening effort (Alhanbali et al., 2019; Francis and Love, 2020; Shields et al., 2023). Therefore, future studies with bigger sample size should explore and add mediating factors (for instance, level of engagement, likelihood of giving up, socio-economic status) to disentangle the multidimensionality of this physiological measure. QoL also related to the quadratic terms of individual PPD psychometric curves in the same direction as clinical word recognition (i.e., participants with higher clinical word and higher QoL had similar PPD psychometric curves than their counterparts) but did not pass Bonferroni correction. It seems that QoL might affect individual variabilities when listeners disengage or perceive the task as unworthy of effort investment. An individual with better status of life quality might maintain motivation or engagement during communication, even when the auditory scene is challenging. Factors that could affect individual decision to disengage need to be explored in future studies.

Some absences of significant relations are surprising. Spectro-temporal auditory acuity has been found to relate to speech recognition in noise and sensitive enough to reliably discriminate CI users (Tamati et al., 2020; Moberly et al., 2021). Therefore, it should not be difficult to assume that this bottom-up factor can also explain individual differences in listening effort and PPD psychometric functions. However, SMRT performance related to the quadratic term of individual PPD psychometric curves in our current study but did not pass the Bonferroni correction. Top-down factors measured in our study relating to working memory and inhibition also did not significantly explain variabilities in average PPD or psychometric functions, despite reports on the possible importance of these cognitive domains on CI users' speech recognition. One explanation to the lack of findings could be due to the mixed results in past studies and variabilities in tests used to measure these cognitive domains and

speech recognition performance (Tao et al., 2014; Skidmore et al., 2020; Tamati et al., 2020; Völter et al., 2021, 2022; Beckers et al., 2023). For instance, different tests have been used to measure working memory, and results in significance level and effect size differ, also depending on which speech recognition outcomes the cognitive tests are related to. Indeed, an increasing number of studies in recent years have started to investigate the impact of cognitive factors on the variability of CI outcomes, but we are still far away from producing a consistent synergy of knowledge. There is still no consensus on the effect size of the relation and what the best tests to quantify these cognitive domains are. And if unexplained variability in CI speech performance is still an issue, the large unexplained variability in CI listening effort that we have observed is not surprising. Another possibility for the lack of significant findings could be that the importance of top-down and bottom-up factors to CI hearing outcomes depends on whether CI user lands in high-or low-performing group. For instance, CI users with high performance using Perpetually Robust English Sentence Test Open-set (PRESTO) also had better SMRT and non-verbal intelligence, but CI users with low PRESTO had more variable individual profiles (Tamati et al., 2020). CI users with low monosyllabic word recognition (<30%) tend to have significant differences in attention and working memory, while CI users with high word recognition (>70%) tend to take advantage of top-down benefits more consistently (Völter et al., 2021). It seems that bottom-up auditory sensitivities are important for CI users to accumulate enough information, but once a certain amount is reached, then the top-down neurocognitive capacity contributes to extra benefit in the hearing outcomes. Therefore, it is likely to have better cognitive capacities that could specifically benefit high performers. In our current study, CI participants were split by word recognition in the median, hence the range of performance was limited (low word group: 50 to 68%; high word group: 73 to 100%). Participants agreeing to attend experiments were already comfortable with doing speech in noise task, a demanding task for CI users, so arguably, our participants were relatively good performers in our CI participants' pool. Therefore, our experiment design is not suitable for comparing peripheral and neurocognitive interactions in bottom-and top-performers. However, our results suggested a similar trend. From low to high word recognition performers, there was a consistent change in PPD psychometric function and SMRT was associated with that consistent change ($r = -0.62$, $p = 0.007$, did not pass Bonferroni correction). When examining whether there is also a group difference in the relation between PPD psychometric function and neurocognitive factors as suggested by past literature, we discovered that it was indeed the case for Stroop and N-back results (Figure 6). It seems that the association between these neurocognitive factors and pupil psychometric function, if exists, might only be exhibited in the high word recognition group. For participants with high word recognition, higher difference between congruent and incongruent performance was associated with smaller quadratic term ($\beta = -0.81$, SE = 0.36, $p = 0.04$), and no significant relationship for participants with low word recognition ($\beta = 0.08$, SE = 0.08, $p = 0.36$); for participants with high word recognition, higher difference between congruent and incongruent performance was associated with bigger average PPD ($\beta = 0.25$, SE = 0.1, $p = 0.03$), and no significant relationship for participants with low word recognition ($\beta = 0.005$, SE = 0.03, $p = 0.82$); For participants with high word recognition, higher d-prime sensitivity was associated with smaller average PPD ($\beta = -0.01$,

SE = 0.01, $p = 0.06$), more so compared to participants with low word recognition ($\beta = -0.07$, SE = 0.02, $p = 0.02$). As a cautious note, these interactions are observed post-hoc, without the explicit control on participants group, and fail strict Bonferroni correction on value of ps. Therefore, future studies should improve on these points to confirm the relationship.
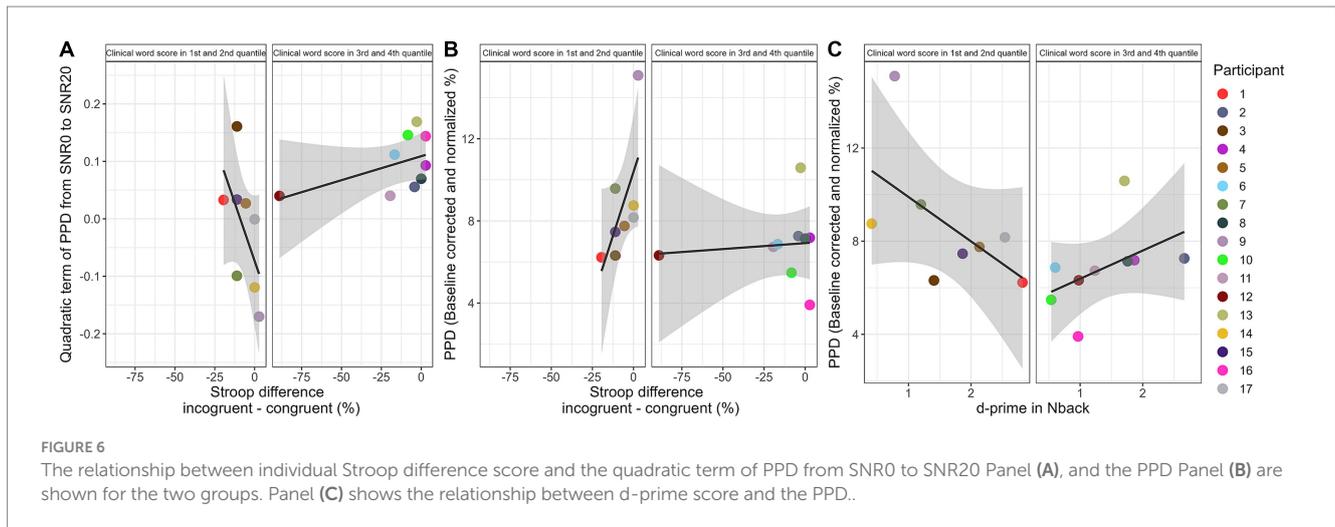
4.3 Limitations

Although our current study sets out with explicit hypothesis on the pupillometry psychometric shape in CI users and its relationship with auditory and neurocognitive factors, there are many interesting post-hoc findings that can be examined and explored. These include, for instance, replicating past findings in the construct relationship across cognitive tests, exploring not main but mediating effects of hearing status and individual differences on hearing outcomes, replicating construct validities of listening effort measures etc. Some of these results are not included in the main paper but can be found in Supplementary Material 2, reported in a descriptive manner. We also did not explore all pupillometry indices, because the aim of the current experiment was to examine whether PPD, the most widely reported listening effort index, can consistently reflect listening effort as in NH and HI listeners. Supplementary Material 2 also provides examination on other pupillary responses (i.e., pupil peak latency and pupil baseline). Future studies with bigger sample size, hence with stronger statistical power, should focus on exploring and validating those relations. Note that past studies have pointed to pupil baseline as a biomarker for variations in arousal and engagement, which could be responsible for the individual differences in pupil psychometric function and the 'tipping point' (Ayasse and Wingfield, 2020; Alhanbali et al., 2021; Micula et al., 2022). Unfortunately, the eye tracker used in our experiment did not measure (in hardware) pupil diameter in absolute unit mm, making the comparison across recording sessions inaccurate. Also, CI participants were prioritized to receive breaks whenever needed, probably affecting arousal and engagement levels independently from the experimental design.

While it is important to replicate past studies and explore new relationships, it is also important to control for family-wise error rate, especially in the context of exploring and discovering new and impactful factors to explain and predict CI hearing outcomes. In fear of type 1 error, i.e., rejecting a null hypothesis that is true in the population, we have applied strict multiple testing value of p corrections, even though this practice is not consistently done in the past literature. Ideally, a structural approach can be more suitable to identify and quantify the importance of individual factors by also considering other factors in the structure. But supporting such models requires bigger dataset and more measurements of different cognitive, listening effort, hearing status and audiology domains (Marsja et al., 2022; Homman et al., 2023).

4.4 Practical considerations in research and clinical applications

This study is motivated by the authors intending to resolve anecdotal reports on using pupillometry with CI users in the clinical settings. Past studies have examined CI users' pupillary response at

The relationship between individual Stroop difference score and the quadratic term of PPD from SNR0 to SNR20 Panel **(A)**, and the PPD Panel **(B)** are shown for the two groups. Panel **(C)** shows the relationship between d-prime score and the PPD..

various points on the psychometric curve (Russo et al., 2020; Stronks et al., 2021; Dingemanse and Goedegebure, 2022), however our study is the first to date to approximate a more complete curve by testing multiple points and considering a wide range of individual differences. The results and their implications need to be carefully digested for research and clinical application.

Firstly, pupillometry remains a reliable method to quantify listening effort, but specifically in a region where speech communication is successful or deemed worthwhile. This is not a foreign concept, considering the guidelines for measuring listening effort in NH and HI listeners (Winn et al., 2018; Keidser et al., 2020). But it is more difficult to implement in CI users, due to their big variabilities in speech outcomes and motivation, as our results have shown. For instance, the assumption that 50% intelligibility is where CI users show PPD peak on their psychometric function might not stand for every CI user. For some CI users, PPD is associated with listening effort and perceived task demand when speech intelligibility is above 60–70%. Apart from careful selection of testing levels, it is still unclear whether and how much engagement or motivation affects PPD psychometric function. However, judging by the difference in the variability of psychometric curves in CI users compared to NH and HI listeners, engagement or motivation probably plays a bigger role in CI users. Therefore, using more ecological speech materials that reflect real communication demands can simulate better the real motivation of CI users to perform successful daily communication (for instance, using well-controlled ecological virtual reality setup as in Pedersen et al. (2023). Arguably, this approach could stabilize the relation between cognitive capacity and effort, and also reveal listening effort in scenarios where it is most relevant (Lunner et al., 2016; Winn et al., 2018; Herrmann and Johnsrude, 2020). Future work on seeking optimal conditions to capture CI users' listening effort in research labs and clinics should consider variabilities in their performance level and motivation.

Secondly, individual differences in CI users are not trivial. Our results join the past literature in showing that top-down and bottom-up individual differences affect CI performance outcomes, but also adds in further complexities by including the dimension of listening effort (Amini et al., 2023; Beckers et al., 2023). Our result also highlights not just the importance of investigating the construct validity across listening effort measures, but also the link between

listening effort and QoL. Relationship between objective outcome measures and Quality of Life might not be straightforward, as shown in our results, so mediating factors need to be taken into consideration to fully understand how to use objective measures such as pupillometry to inform and advise patients' care. It could be beneficial to be aware of the importance of individual variabilities in hearing outcomes at direct points of care (hospitals, audiology clinics, rehabilitation centers and research labs), so that a holistic and responsible database can be built that contains individual factors such as hearing status, neurocognitive functioning, auditory sensitivity, socio-economic status, surgical outcomes etc. (see[2] for an example). Such database should also follow strict privacy and ethics compliance such as General Data Protection Regulation (GDPR). Further guidelines could be helpful to be agreed upon across multiple sites on what are the criteria for the discovery of meaningful individual factors, such as what are acceptable type 1 and type 2 errors rates, what are clinically relevant effect sizes, how to include mediating factors, how to discover bias in database and site experimental practices, etc.

5 Conclusion

To summarize, our study replicates, in CI users, the well-validated NH and HI listeners' inverse U-shape psychometric function between pupillary response and speech intelligibility across a wide range of SNR levels. This psychometric relationship supports our interpretation of pupillary response as a biomarker of listening effort, hence important for both research and clinical practices. Our results suggest that big variabilities in individual psychometric relationships disturb the construction of a consistent psychometric function at the group levl. However, some individual factors relating to CI users' speech outcomes, quality of life and neurocognitive functioning can predict individual psychometric shapes. Future studies should examine the possible effect of motivation or engagement in modulating CI users' occurrence of 'tipping point' on their psychometric functions. It is likely that a new biomarker indicating when CI users 'give up' could

---

2 https://plataforma.innovacionsalud.org

be an important holistic benchmark for future CI strategy and rehabilitation innovations to validate their effects in improving CI users' engagement in speech communication.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Hospital Universitario Virgen Macarena in Seville, Spain (PEIBA#19/2019). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

YZ and FP are employees of Oticon Medical, manufacturer of the cochlear implants used by the patients in this study and receive salaries as part of their employment. MC-L's research is funded by a Torres Quevedo grant from the Spanish Ministry of Science and Innovation (PTQ 2019-010594/AEI/10.13039/501100011033) and co-funded by Oticon Medical.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could cause a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1307777/full#supplementary-material

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Scand. Audiol.* 47, S53–S71. doi: 10.1080/14992020802301142

Alhanbali, S., Millman, R. E., Munro, K. J., and Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear Hear.* 40, 1084–1097. doi: 10.1097/AUD.0000000000000697

Alhanbali, S., Munro, K. J., Dawes, P., Carolan, P. J., and Millman, R. E. (2021). Dimensions of self-reported listening effort and fatigue on a digits-in-noise task, and association with baseline pupil size and performance accuracy 60, 762–772. doi: 10.1080/14992027.2020.1853262

Amini, A. E., et al. (2023). Emerging relations among cognitive constructs and Cochlear implant outcomes: a systematic review and Meta-analysis. *Otolaryngology*. doi: 10.1002/OHN.344

Aronoff, J. M., and Landsberger, D. M. (2013). The development of a modified spectral ripple test. *J. Acoust. Soc. Am.* 134, EL217–EL222. doi: 10.1121/1.4813802

Ayasse, N. D., and Wingfield, A. (2020). Anticipatory baseline pupil diameter is sensitive to differences in hearing thresholds. *Front. Psychol.* 10:504013. doi: 10.3389/FPSYG.2019.02947/BIBTEX

Bates, D., Bolker, B., Walker, S., and Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beckers, L., Tromp, N., Philips, B., Mylanus, E., and Huinck, W. (2023). Exploring neurocognitive factors and brain activation in adult cochlear implant recipients associated with speech perception outcomes—a scoping review. *Front. Neurosci.* 17:1046669. doi: 10.3389/fnins.2023.1046669

Bergeron, F., and Hotton, M. (2016). Perception in noise with the Digisonic SP cochlear implant: clinical trial of Saphyr processor's upgraded signal processing. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 133, S4–S6. doi: 10.1016/J.ANORL.2016.04.019

Besser, J., Zekveld, A. A., Kramer, S. E., Festen, J. M., and Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—a review. *Trends Amplif.* 17, 75–93. doi: 10.1177/1084713813495459

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., et al. (2012). Factors affecting auditory performance of Postlinguistically deaf adults using Cochlear implants: an update with 2251 patients. *Audiol. Neurotol.* 18, 36–47. doi: 10.1159/000343189

Boisvert, I., Reis, M., Au, A., Cowan, R., and Dowell, R. C. (2020). Cochlear implantation outcomes in adults: a scoping review. *PLoS One* 15:e0232421. doi: 10.1371/JOURNAL.PONE.0232421

Bönitz, H., Lunner, T., Finke, M., Fiedler, L., Lyxell, B., Riis, S. K., et al. (2021). How do we allocate our resources when listening and memorizing speech in noise? A Pupillometry study. *J. Am. Aud. Soc.* 42, 846–859. doi: 10.1097/AUD.0000000000001002

Bozorg-Grayeli, A., Guevara, N., Bebear, J. P., Ardoint, M., Saaï, S., Hoen, M., et al. (2016). Clinical evaluation of the xDP output compression strategy for cochlear implants. *Eur. Arch. Otorhinolaryngol.* 273, 2363–2371. doi: 10.1007/s00405-015-3796-1

Brehm, J. W., and Self, E. A. (1989). The intensity of motivation. *Annu. Rev. Dev. Psychol.* 40, 109–131. doi: 10.1146/annurev.ps.40.020189.000545

Burg, E. A., et al. (2021). Systematic comparison of trial exclusion criteria for Pupillometry data analysis in individuals with single-sided deafness and Normal hearing. *Trends Hear.* 25, 1–17. doi: 10.1177/23312165211013256

Carolan, P. J., Munro, K. J., Millman, R. E., and Millman, R. E. (2022). Quantifying the effects of motivation on listening effort: a systematic review and Meta-analysis. *Trends Hear.* 26:233121652110599. doi: 10.1177/23312165211059982

Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the raven progressive matrices test. *Psychol. Rev.* 97, 404–431. doi: 10.1037/0033-295X.97.3.404

Cordeiro, B. B., et al. (2021). Adult users of the Oticon medical neuro Cochlear implant system benefit from beamforming in the high frequencies. *Audiol. Res.* 11, 179–191. doi: 10.3390/AUDIOLRES11020016

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verbal. Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

de Cárdenas, M. R., and Marrero, V. (1994). *Cuaderno de logoaudiometría: Guía de referencia rápida.* Madrid, Spain: Universidad Nacional de Educación a Distancia.

Dingemanse, G., and Goedegebure, A. (2022). Listening effort in Cochlear implant users: the effect of speech intelligibility, noise reduction processing, and working memory capacity on the pupil dilation response. *J. Speech Lang. Hear. Res.* 65, 392–404. doi: 10.1044/2021_JSLHR-21-00230

DiNino, M., and Arenberg, J. G. (2018). Age-related performance on vowel identification and the spectral-temporally modulated ripple test in children with Normal hearing and with Cochlear implants. *Trends Hear.* 22, 1–20. doi: 10.1177/2331216518770959

Dorman, M. F., and Gifford, R. H. (2017). Speech understanding in complex listening environments by listeners fit with Cochlear implants. *J. Speech Hear. Res.* 60, 3019–3026. doi: 10.1044/2017_JSLHR-H-17-0035

Francis, A. L., and Love, J. (2020). Listening effort: are we measuring cognition or affect, or both? *Wiley Interdiscip. Rev. Cogn. Sci.* 11:e1514. doi: 10.1002/WCS.1514

Füllgrabe, C., and Rosen, S. (2016). Investigating the role of working memory in speech-in-noise identification for listeners with Normal hearing. *Physiol. Psychoacoust. Cogn. Norm. Impair. Hear.* 894, 29–36. doi: 10.1007/978-3-319-25474-6_4

Gevins, A., and Cutillo, B. (1993). Spatiotemporal dynamics of component processes in human working memory. *Electroencephalogr. Clin. Neurophysiol.* 87, 128–143. doi: 10.1016/0013-4694(93)90119-G

Goudey, B., Plant, K., Kiral, I., Jimeno-Yepes, A., Swan, A., Gambhir, M., et al. (2021). A MultiCenter analysis of factors associated with hearing outcome for 2,735 adults with Cochlear implants. *Trends Hear.* 25:110375. doi: 10.1177/23312165211037525

Guevara, N., Bozorg-Grayeli, A., Bebear, J. P., Ardoint, M., Saaï, S., Gnansia, D., et al. (2016). The voice track multiband single-channel modified wiener-filter noise reduction system for cochlear implants: patients' outcomes and subjective appraisal. *Scand. Audiol.* 55, 431–438. doi: 10.3109/14992027.2016.1172267

Haatveit, B. C., Sundet, K., Hugdahl, K., Ueland, T., Melle, I., and Andreassen, O. A. (2010). The validity of d prime as a working memory index: results from the Bergen n-back task. *J. Clin. Exp. Neuropsychol.* 32, 871–880. doi: 10.1080/13803391003596421

Harris, A. M., McMillan, J., Listyg, B., Matzen, L., and Carter, N. (2020). Measuring intelligence with the Sandia matrices: psychometric review and recommendations for free raven-like item sets. *Person. Assess. Decisions* 6:6. doi: 10.25035/pad.2020.03.006

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Human Mental Workload* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Herrmann, B., and Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hear. Res.* 397:108016. doi: 10.1016/J.HEARES.2020.108016

Hinderink, J. B., Krabbe, P. F. M., and Van Den Broek, P. (2000). Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen Cochlear implant questionnaire. *Otolaryngol. Head Neck Surg.* 123, 756–765. doi: 10.1067/MHN.2000.108203

Holden, L. K., Finley, C. C., Firszt, J. B., Holden, T. A., Brenner, C., Potts, L. G., et al. (2013). Factors affecting open-set word recognition in adults with Cochlear implants. *J. Am. Aud. Soc.* 34, 342–360. doi: 10.1097/AUD.0B013E3182741AA7

Homman, L., Danielsson, H., and Rönnberg, J. (2023). A structural equation mediation model captures the predictions amongst the parameters of the ease of language understanding model concentration, task engagement and distraction view project alternative communication view project a structural equation mediation model captures the predictions amongst the parameters of the ease of language understanding model' 14, 1–20. doi: 10.3389/fpsyg.2023.1015227

Hughes, K. C., and Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int.* 14, 121–129. doi: 10.1179/1754762812Y.0000000009

Kassner, M., Patera, W., and Bulling, A. (2014) 'Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction', UbiComp 2014 - Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Association for Computing Machinery, Inc, 1151–1160.

Keidser, G., Naylor, G., Brungart, D. S., Caduff, A., Campos, J., Carlile, S., et al. (2020). The quest for ecological validity in hearing science: what it is, why it matters, and how to advance it. *J. Am. Aud. Soc.* 41, 5S–19S. doi: 10.1097/AUD.0000000000000944

Kestens, K., Degeest, S., and Keppler, H. (2021). The effect of cognition on the aided benefit in terms of speech understanding and listening effort obtained with digital hearing aids: a systematic review. *Am. J. Audiol.* 30, 190–210. doi: 10.1044/2020_AJA-20-00019

Koelewijn, T., Shinn-Cunningham, B. G., Zekveld, A. A., and Kramer, S. E. (2014). The pupil response is sensitive to divided attention during speech processing. *Hear. Res.* 312, 114–120. doi: 10.1016/J.HEARES.2014.03.010

Koelewijn, T., Zekveld, A. A., Festen, J. M., and Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.* 33, 291–300. doi: 10.1097/AUD.0B013E3182310019

Kothe, C., Medine, D., Boulay, C., Grivich, M., and Stenner, T. (2012) Lab streaming layer (LSL)-a software framework for synchronizing a large array of data collection and stimulation devices, The Swartz Center for Computational Neuroscience Software. San Diego, USA: Available at: https://code.google.com/archive/p/labstreaminglayer/. (Accessed 29 February 2020)

Kramer, S. E., Kapteyn, T. S., Festen, J. M., and Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Int. J. Audiol.* 36, 155–164. doi: 10.3109/00206099709071969

Lawler, M., Yu, J., and Aronoff, J. M. (2017). Comparison of the spectral-temporally modulated ripple test with the Arizona biomedical institute sentence test in Cochlear implant users. *J. Am. Aud. Soc.* 38, 760–766. doi: 10.1097/AUD.0000000000000496

Lazard, D. S., Vincent, C., Venail, F., van de Heyning, P., Truy, E., Sterkers, O., et al. (2012). Pre-, per-and postoperative factors affecting performance of Postlinguistically deaf adults using Cochlear implants: a new conceptual model over time. *PLOS ONE* 7:e48739. doi: 10.1371/JOURNAL.PONE.0048739

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/J.JESP.2013.03.013

Lunner, T., Rudner, M., Rosenbom, T., Ågren, J., and Ng, E. H. (2016). Using speech recall in hearing aid fitting and outcome evaluation under ecological test conditions. *J. Am. Aud. Soc.* 37 Suppl 1, 145S–154S. doi: 10.1097/AUD.0000000000000294

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. doi: 10.21105/joss.00470,

Marsja, E., Stenbäck, V., Moradi, S., Danielsson, H., and Rönnberg, J. (2022). Is having hearing loss fundamentally different? Multigroup structural equation modeling of the effect of cognitive functioning on speech identification. *J. Am. Aud. Soc.* 43, 1437–1446. doi: 10.1097/AUD.0000000000001196

Mattingly, J. K., Castellanos, I., and Moberly, A. C. (2018). 'Nonverbal reasoning as a contributor to sentence recognition outcomes in adults with Cochlear implants', otology & neurotology: official publication of the American Otological Society, American Neurotology Society [and] European academy of otology and Neurotology. *Am. J. Otol.* 39:e956, –e963. doi: 10.1097/MAO.0000000000001998

Matzen, L. E., Benz, Z. O., Dixon, K. R., Posey, J., Kroger, J. K., and Speed, A. E. (2010). Recreating raven's: software for systematically generating large numbers of raven-like matrix problems with normed properties. *Behav. Res. Methods Instrum. Comput.* 42, 525–541. doi: 10.3758/BRM.42.2.525

McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., and Munro, K. J. (2017). Measuring listening-related effort and fatigue in school-aged children using pupillometry. *J. Exp. Child Psychol.* 161, 95–112. doi: 10.1016/J.JECP.2017.04.006

Micula, A., Rönnberg, J., Fiedler, L., Wendt, D., Jørgensen, M. C., Larsen, D. K., et al. (2021). The effects of task difficulty predictability and noise reduction on recall performance and pupil dilation responses. *J. Am. Aud. Soc.* 42, 1668–1679. doi: 10.1097/AUD.0000000000001053

Micula, A., Rönnberg, J., Zhang, Y., and Ng, E. H. N. (2022). A decrease in physiological arousal accompanied by stable behavioral performance reflects task habituation. *Front. Neurosci.* 16:876807. doi: 10.3389/fnins.2022.876807

Mirman, D. (2017) Growth curve analysis and visualization using R. Available at:https://books.google.com/books?hl=en&dr=&id=vkQPEAAAQBAJ&oi=fnd&pg=PP

1&dq=Mirman+D.+Growth+curve+analysis+and+visualization+using+R.+CRC+Press+Boca+Raton,+FL%3B+2014.&ots=ODNn67qxq3&sig=De_Cxvg6blfDjHvOb2lQ6sJYzqM (Accessed: 26 July 2023).

Moberly, A. C., Harris, M. S., Boyce, L., and Nittrouer, S. (2017a). Speech recognition in adults with Cochlear implants: the effects of working memory, phonological sensitivity, and aging. *J. Speech Lang. Hear. Res.* 60, 1046–1061. doi: 10.1044/2016_JSLHR-H-16-0119

Moberly, A. C., Harris, M. S., Boyce, L., Vasil, K., Wucinich, T., Pisoni, D. B., et al. (2018). Relating quality of life to outcomes and predictors in adult cochlear implant users: are we measuring the right things? *Laryngoscope Invest. Otolaryngol.* 128, 959–966. doi: 10.1002/LARY.26791

Moberly, A. C., Houston, D. M., Harris, M. S., Adunka, O. F., and Castellanos, I. (2017b). Verbal working memory and inhibition-concentration in adults with cochlear implants. *Laryngoscope Invest. Otolaryngol.* 2, 254–261. doi: 10.1002/LIO2.90

Moberly, A. C., Lewis, J. H., Vasil, K. J., Ray, C., and Tamati, T. N. (2021). Bottom-up signal quality impacts the role of top-down cognitive-linguistic processing during speech recognition by adults with Cochlear implants. *Otol. Neurotol.* 42, S33–S41. doi: 10.1097/MAO.0000000000003377

Moberly, A. C., Mattingly, J. K., and Castellanos, I. (2019). How does nonverbal reasoning affect sentence recognition in adults with Cochlear implants and Normal-hearing peers? *Audiol. Neurotol.* 24, 127–138. doi: 10.1159/000500699

Murphy, P. R., Robertson, I. H., Balsters, J. H., and O'connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology* 48, 1532–1543. doi: 10.1111/J.1469-8986.2011.01226.X

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., and Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. *Scand. Audiol.* 48, 684–691. doi: 10.1080/14992020902962421

Ohlenforst, B., Wendt, D., Kramer, S. E., Naylor, G., Zekveld, A. A., and Lunner, T. (2018). Impact of SNR, masker type and noise reduction processing on sentence recognition performance and listening effort as indicated by the pupil dilation response. *Hear. Res.* 365, 90–99. doi: 10.1016/J.HEARES.2018.05.003

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., et al. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear. Res.* 351, 68–79. doi: 10.1016/J.HEARES.2017.05.012

Pedersen, R. L., et al. (2023). Virtual-reality-based research in hearing science: a platforming approach. *J. Audio Eng. Soc.* 71, 374–389. doi: 10.17743/JAES.2022.0083

Perea Pérez, F., Hartley, D. E. H., Kitterick, P. T., Zekveld, A. A., Naylor, G., and Wiggins, I. M. (2023). Listening efficiency in adult cochlear-implant users compared with normally-hearing controls at ecologically relevant signal-to-noise ratios. *Front. Hum. Neurosci.* 17, 1214485. doi: 10.3389/fnhum.2023.1214485

Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., and Corts, D. (2017). Listening effort measured in adults with normal hearing and cochlear implants. *J. Am. Acad. Audiol.* 28, 685–697. doi: 10.3766/jaaa.16014

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *J. Am. Aud. Soc.* 37 Suppl 1, 5S–27S. doi: 10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Mick, P., and Reed, M. (2015). Hearing, cognition, and healthy aging: social and public health implications of the links between age-related declines in hearing and cognition. *Seminars Hear.* 36, 122–139. doi: 10.1055/S-0035-1555116/ID/JR00675-28/BIB

Pisoni, D. B. (2000). *Cognitive factors and Cochlear implants: Some thoughts on perception, learning, and memory in speech perception.* Miyamoto, 21, 70–78. doi: org/10.1097/00003446-200002000-00010.

Pisoni, D. B., and Cleary, M. (2003). Measures of working memory span and verbal rehearsal Speed in deaf children after Cochlear implantation. *J. Am. Aud. Soc.* 24, 106S–120S. doi: 10.1097/01.AUD.0000051692.05140.8E

Rapport, F., et al. (2020). 'Adults' cochlear implant journeys through care: a qualitative study. *Bio Med. Central* 20, 1–9. doi: 10.1186/S12913-020-05334-Y/TABLES/1

Richter, M. (2013). A closer look into the multi-layer structure of motivational intensity theory. *Soc. Pers. Psych.* 7, 1–12. doi: 10.1111/SPC3.12007

Richter, M. (2016). The moderating effect of success importance on the relationship between listening demand and listening effort. *J. Am. Aud. Soc.* 37, 111S–117S. doi: 10.1097/AUD.0000000000000295

Rudner, M., Rönnberg, J., and Lunner, T. (2011). Working memory supports listening in noise for persons with hearing impairment. *J. Am. Acad. Audiol.* 22, 156–167. doi: 10.3766/JAAA.22.3.4/BIB

Russo, F. Y., et al. (2020). Pupillometry assessment of speech recognition and listening experience in adult Cochlear implant patients. *Front. Neurosci.* 14:1115. doi: 10.3389/FNINS.2020.556675/BIBTEX

Shields, C., Sladen, M., Bruce, I. A., Kluk, K., and Nichani, J. (2023). Exploring the correlations between measures of listening effort in adults and children: a systematic review with narrative synthesis. *Trends Hear.* 27:11371. doi: 10.1177/23312165221137116

Shukla, A., Harper, M., Pedersen, E., Goman, A., Suen, J. J., Price, C., et al. (2020). Hearing loss, loneliness, and social isolation: a systematic review. *Otolaryngol. Head Neck Surg.* 162, 622–633. doi: 10.1177/0194599820910377

Skidmore, J. A., Vasil, K. J., He, S., and Moberly, A. C. (2020). Explaining speech recognition and quality of life outcomes in adult Cochlear implant users: complementary contributions of demographic, sensory, and cognitive factors. *Otol. Neurotol.* 41:e795, –e803. doi: 10.1097/MAO.0000000000002682

Smeds, K., Wolters, F., and Rung, M. (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *J. Am. Acad. Audiol.* 26, 183–196. doi: 10.3766/jaaa.26.2.7

Strand, J. F., Brown, V. A., Merchant, M. B., Brown, H. E., and Smith, J. (2018). Measuring listening effort: convergent validity, sensitivity, and links with cognitive and personality measures. *J. Speech Lang. Hear. Res.* 61, 1463–1486. doi: 10.1044/2018_JSLHR-H-17-0257

Stronks, H. C., Apperloo, E., Koning, R., Briaire, J. J., and Frijns, J. H. M. (2021). SoftVoice improves speech recognition and reduces listening effort in Cochlear implant users. *Ear Hear.* 42, 381–392. doi: 10.1097/AUD.0000000000000928

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/H0054651

Sulas, E., et al. (2022). Streamlining experiment design in cognitive hearing science using OpenSesame. *Behav. Res. Methods Springer* 1, 1–15. doi: 10.3758/S13428-022-01886-5/FIGURES/14

Tamati, T. N., and Moberly, A. C. (2022). Talker adaptation and lexical difficulty impact word recognition in adults with Cochlear implants. *Audiol. Neurotol.* 27, 260–270. doi: 10.1159/000518643

Tamati, T. N., Ray, C., Vasil, K. J., Pisoni, D. B., and Moberly, A. C. (2020). High- and low-performing adult Cochlear implant users on high-variability sentence recognition: differences in auditory spectral resolution and neurocognitive functioning. *J. Am. Acad. Audiol.* 31, 324–335. doi: 10.3766/jaaa.18106

Tao, D., Deng, R., Jiang, Y., Galvin, J. J., Fu, Q. J., and Chen, B. (2014). Contribution of auditory working memory to speech understanding in mandarin-speaking Cochlear implant users. *PLoS One* 9:e99096. doi: 10.1371/JOURNAL.PONE.0099096

Veneman, C. E., Gordon-Salant, S., Matthews, L. J., and Dubno, J. R. (2013). Age and measurement time-of-day effects on speech recognition in noise. *Ear Hear.* 34, 288–299. doi: 10.1097/AUD.0B013E31826D0B81

Völter, C., Götze, L., Bajewski, M., Dazert, S., and Thomas, J. P. (2022). Cognition and cognitive Reserve in Cochlear Implant Recipients. *Front. Aging Neurosci.* 14:214. doi: 10.3389/FNAGI.2022.838214

Völter, C., Oberländer, K., Carroll, R., Dazert, S., Lentz, B., Martin, R., et al. (2021). Nonauditory functions in low-performing adult Cochlear implant users. *Am. J. Otol.* 42, E543–E551. doi: 10.1097/MAO.0000000000003033

Wagner, A. E., et al. (2019). Individual variations in effort: assessing Pupillometry for the hearing impaired. *Trends Hear.* 23, 1–18. doi: 10.1177/2331216519845596

Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B., et al. (2018). Relations between Self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task. *Ear Hear.* 39, 573–582. doi: 10.1097/AUD.0000000000000512

Wang, Y., Zekveld, A. A., Naylor, G., Ohlenforst, B., Jansma, E. P., Lorens, A., et al. (2016). Parasympathetic nervous system dysfunction, as identified by pupil light reflex, and its possible connection to hearing impairment. *PLoS One* 11:e0153566. doi: 10.1371/JOURNAL.PONE.0153566

Wendt, D., Hietkamp, R. K., and Lunner, T. (2017). Impact of noise and noise reduction on processing effort: a Pupillometry study. *J. Am. Aud. Soc.* 38, 690–700. doi: 10.1097/AUD.0000000000000454

Wickham, H. (2016). '*Data Analysis*'. Cham: Springer.

Winn, M. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and Cochlear implants. *Trends Hear.* 20, 1–17. doi: 10.1177/2331216516669723

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear.* 36:e153, –e165. doi: 10.1097/AUD.0000000000000145

Winn, M. B., and Teece, K. H. (2022). Effortful listening despite correct responses: the cost of mental repair in sentence recognition by listeners with Cochlear implants. *American Speech-Language-Hearing Association* 65, 3966–3980. doi: 10.1044/2022_JSLHR-21-00631

Winn, M. B., et al. (2018). Best practices and advice for using Pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear.* 22, 1–32. doi: 10.1177/2331216518800869

Zekveld, A. A., Koelewijn, T., and Kramer, S. E. (2018). The pupil dilation response to auditory stimuli: current state of knowledge. *Trends Hear.* 22, 1–25. doi: 10.1177/2331216518777174

Zekveld, A. A., and Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology* 51, 277–284. doi: 10.1111/PSYP.12151

Zekveld, A. A., Kramer, S. E., Rönnberg, J., and Rudner, M. (2019). In a concurrent memory and auditory perception task, the pupil dilation response is more sensitive to memory load than to auditory stimulus characteristics. *J. Am. Aud. Soc.* 40, 272–286. doi: 10.1097/AUD.0000000000000612

Zhang, Y., Deroche, M., and Lehmann, A. (2021). Disentangling listening effort and memory load beyond behavioural evidence: pupillary response to listening effort during a concurrent memory task. *PLoS One* 16:e0233251. doi: 10.1371/JOURNAL.PONE.0233251

Zhang, Y., Malaval, F., Lehmann, A., and Deroche, M. L. D. (2022). Luminance effects on pupil dilation in speech-in-noise recognition. *PLoS One* 17:e0278506. doi: 10.1371/JOURNAL.PONE.0278506