scientific reports

Check for updates

OPEN

Sensory representations and pupil-indexed listening effort provide complementary contributions to multi-talker speech intelligibility

Jacie R. McHaney[1,5], Kenneth E. Hancock[2,3], Daniel B. Polley[2,3] & Aravindakshan Parthasarathy[1,4]✉

Multi-talker speech intelligibility requires successful separation of the target speech from background speech. Successful speech segregation relies on bottom-up neural coding fidelity of sensory information and top-down effortful listening. Here, we studied the interaction between temporal processing measured using Envelope Following Responses (EFRs) to amplitude modulated tones, and pupil-indexed listening effort, as it related to performance on the Quick Speech-in-Noise (QuickSIN) test in normal-hearing adults. Listening effort increased at the more difficult signal-to-noise ratios, but speech intelligibility only decreased at the hardest signal-to-noise ratio. Pupil-indexed listening effort and EFRs did not independently relate to QuickSIN performance. However, the combined effects of both EFRs and listening effort explained significant variance in QuickSIN performance. Our results suggest a synergistic interaction between sensory coding and listening effort as it relates to multi-talker speech intelligibility. These findings can inform the development of next-generation multi-dimensional approaches for testing speech intelligibility deficits in listeners with normal-hearing.

Keywords Speech-in-noise, Listening effort, Envelope following responses, Pupillometry, Frequency following responses, Cognitive load

Everyday listening in multi-talker environments involves a complex interplay between the neural encoding of acoustic information, and the top-down influences of cognitive factors that contribute to effortful listening. Current clinical assessments of hearing impairments place heavy emphasis on peripheral auditory function and far less on the sequelae of neural coding that follows cochlear transduction and hence are inadequate in capturing this inherently multi-dimensional process. The standard audiological battery is not sensitive enough to capture listening difficulties that are reported by 5–10% of patients who seek help at the audiology clinic, yet present with normal hearing thresholds[1–3].

Deficiencies in neural encoding of afferent sensory information within the auditory pathway reflects a complex mixture of degeneration and compensation at successive stations of auditory processing. Sensory degenerations, such as the loss of outer hair cells or strial function, can manifest as an increase in hearing thresholds. However, cochlear neural deafferentation caused by the loss of inner hair cells, spiral ganglion cells or the synapses between the inner hair cells and the auditory nerve are prevalent but remain 'hidden' to current clinical tests[4–9]. Peripheral deafferentation is often accompanied by compensatory neural plasticity, or a relative increase of activity in central auditory structures[10–13]. Although this increased central 'gain' may benefit listening in quiet environments, it is likely maladaptive for listening in noise[14].

In animal models, changes in peripheral neural encoding and the resultant compensatory gain can be measured by directly assessing relative neural activity in ascending auditory structures[10,13–15]. However, in humans, indirect measurements of central gain are typically acquired using auditory evoked potentials - ensembles of neural activity to sound recorded at the scalp[16,17]. One such auditory evoked potential is the envelope following response (EFR). EFRs reflect steady-state potentials evoked by the neural synchronization, or phase-locking, to an auditory stimulus amplitude envelope[18,19]. The phase-locking capabilities of neurons are

[1]Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA 15260, USA. [2]Deparment of Otolaryngology – Head and Neck Surgery, Harvard Medical School, Boston, MA 02115, USA. [3]Eaton-Peabody Laboratories, Massachusetts Eye and Ear, Boston, MA 02114, USA. [4]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA 15260, USA. [5]Present address: Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208, USA. ✉email: aravind_partha@pitt.edu

biophysically constrained, such that the upper limit of phase-locking decreases along the ascending auditory pathway. Specifically, the upper limit of phase-locking at cortical regions of the auditory pathway are ~80 Hz, while phase-locking limits of the auditory nerve exceed ~1000 Hz (see for review[20]). By exploiting these divergent phase-locking limits of the auditory pathway, EFRs to stimuli with different temporal modulations of the amplitude envelope can emphasize cortical, midbrain, and brainstem sources[8,18,19,21–23]. Hence, by comparing EFRs to fast temporal modulation rates and EFRs to slower temporal modulation rates, one can ideally obtain a picture of auditory temporal processing abilities along the entire auditory neuraxis.

In addition to the fidelity of sensory encoding, multi-talker speech intelligibility involves the recruitment of additional cognitive resources that are diverted to assist with listening[24]. This intentional reallocation of cognitive resources, broadly referred to as listening effort, can be indexed using task-related changes in pupil diameter[24–26]. Increases in task-related pupil size have been associated with cognitive processes, task difficulty, arousal, and speech intelligibility[27–33]. In a prior study, we measured pupil-indexed listening effort while participants identified spoken streams of monosyllabic digits, which are devoid of linguistic context, in the presence of multiple competing digit streams[2]. Changes in pupil diameter during listening were modulated by task difficulty and related to behavioral outcomes. However, it remains unclear the extent to which increases in listening effort change with added linguistic context, and the extent to which these top-down cognitive processes interact with sensory encoding in the auditory pathway.

In the current study, we demonstrate that EFRs can be reliably measured to a variety of modulation frequencies to emphasize complementary neural generators along the ascending auditory pathway in a cohort of young adults with normal hearing thresholds. Further, pupil-indexed listening effort was modulated by task difficulty in participants performing the Quick Speech in Noise test (QuickSIN[34]), a clinically relevant multitasker speech intelligibility task with moderate linguistic and contextual cues. Finally, using a multivariate regression model, we show that bottom-up sensory coding and top-down listening effort provide complementary contributions to multi-talker speech intelligibility.

Results

EFRs can be reliably measured to assess auditory temporal processing for modulation frequencies up to 1024 Hz in humans

We first examined the extent to which EFRs could be reliably recorded for the range of modulation frequencies used in this study. EFR amplitudes exhibit a low-pass shape as modulation frequency increases, such that the temporal modulation transfer function decreases logarithmically. Yet, studies in animal models suggest that EFRs to modulation frequencies in the 500–1000 Hz AM region can still be recorded reliably above the noise floor[8,19]. Here, we sought to determine if the same was true for humans with our recording setup. Figure 1 shows the AM stimuli (Fig. 1A), the grand averaged EFR responses in the time domain (Fig. 1B) and the grand averaged FFT spectra of these EFRs in the frequency domain (Fig. 1C). All four modulation frequencies used in this study exhibited robust EFRs, observed both in the time domain, and more clearly, in the frequency domain. In the frequency domain the peaks at modulation frequency were significantly above the noise floor. A quantification of the signal to noise ratio (SNR) revealed an average SNR of approximately 6 dB *(M* = 6.252 dB, *SD* = 0.850 dB*)* across all modulation frequencies tested (Fig. 1D), suggesting that we can reliably record EFRs within this range in our participant population.

Multi-channel acquisition suggests that EFRs can be utilized to evaluate the relative neural activity from peripheral and central auditory regions

We then sought to confirm that EFRs to various modulation frequencies do in fact emphasize complementary neural generators along the auditory pathway, by leveraging our multichannel approach and comparing EFRs across various electrode montages. The schematic for the hypothesized rationale is displayed in Fig. 2A. Neural generators are indicated by circles whose sizes correspond to size of the auditory nuclei, with cortical generators having the largest size (orange), and peripheral generators having the smallest (purple). The three electrode montages we compared EFRs across were: (1) Fz to ipsilateral tiptrode placed in the stimulated ear (Fz-R), (2) ipsilateral tiptrode to the contralateral tiptrode placed in the unstimulated ear (L-R), and (3) Fz to contralateral tiptrode (Fz-L). The Fz-R montage should capture contributions from all auditory generators along the ascending pathway[23,35]. The L-R montage should capture peripheral generators while de-emphasizing cortical generators due to the distance the volume conducted signal needs to pass through to be captured by the electrode[35,36]. The Fz-L montage should capture central generators while de-emphasizing ipsilateral (to the sound presentation) peripheral generators for the same reason above. This hypothesis was supported by the comparative EFRs shown in Fig. 2B, plotted both as absolute amplitude (left column) and relative change in amplitude compared to the Fz-R condition (right column).

EFRs to 40 Hz AM had the highest response amplitudes in the Fz-R montage, relative to Fz-L ($\beta = -0.047$, $t = -2.785$, $p = .009$) and L-R montages ($\beta = -0.081$, $t = -4.769$, $p < .001$). Response amplitudes significantly decreased by about 30% in the Fz-L montage ($M = 0.110$, $SD = 0.061$) relative to the Fz-R montage ($M = 0.157$, $SD = 0.083$), presumably due to the loss of contributions from subcortical areas. Response amplitudes were further reduced by approximately 40% in the L-R horizontal montage ($M = 0.076$, $SD = 0.044$) relative to the Fz-R montage. At 110 Hz AM rate, EFR amplitudes at Fz-R did not significantly differ from EFR amplitudes in the L-R montage ($\beta = 0.009$, $t = 1.329$, $p = .194$) nor Fz-L montage ($\beta = 0.004$, $t = 0.558$, $p = .581$). The similar EFR amplitudes at 110 Hz across montages was consistent with the idea that ~110 Hz AM reflects a mixture of contributions from cortical and subcortical generators[37,38], and as such would not differ significantly between electrode montages. EFRs to faster AM rates (i.e., 512 and 1024 Hz) demonstrated changes in amplitudes consistent with generators originating from more peripheral neural sources. EFR amplitudes for 512 Hz AM rate were significantly higher in the L-R montage ($\beta = 0.006$, $t = 3.314$, $p = .003$) relative to the Fz-R montage, though

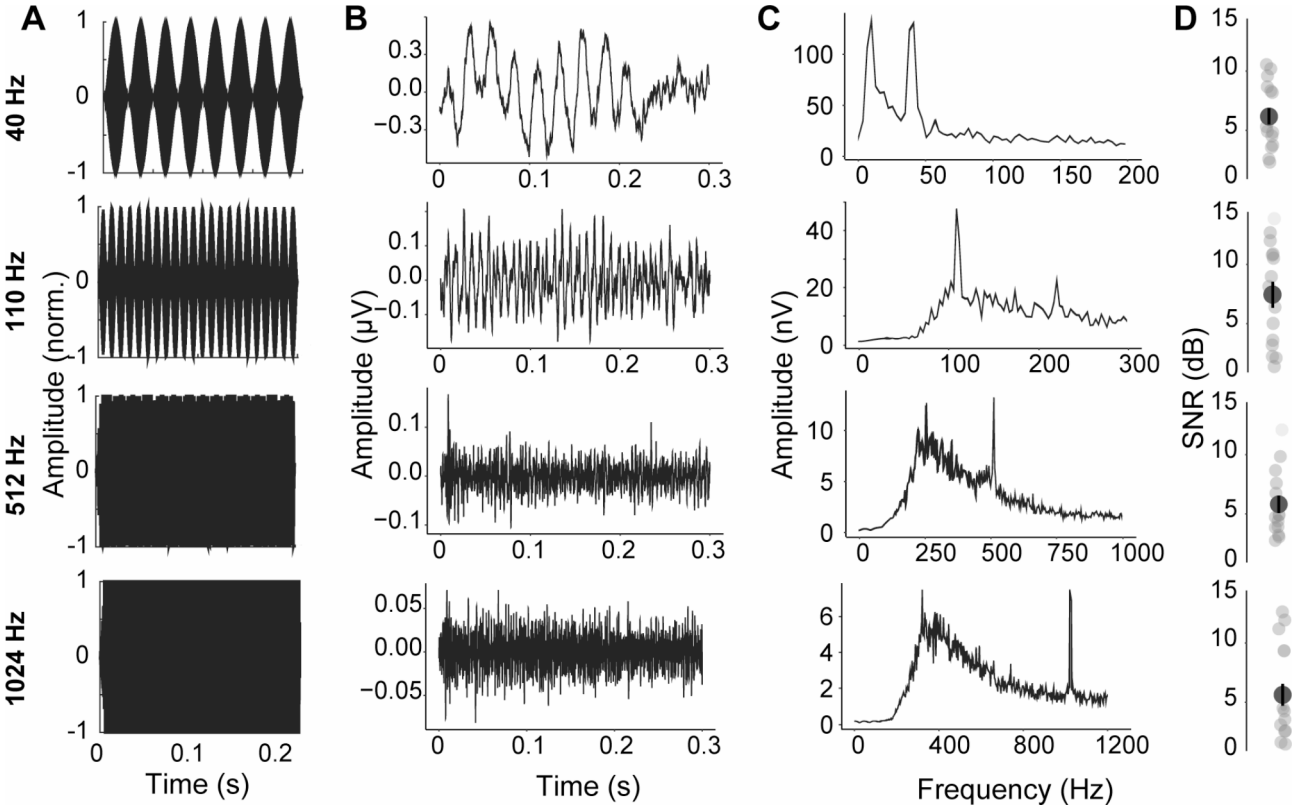

Fig. 1. EFRs can be reliably recorded up to modulation frequencies of 1024 Hz. **(A)** Time domain stimulus waveforms of the AM tones used in this study. The carrier used was 3 kHz, the modulation depth was 100% and the AM rates were 40 Hz, 110 Hz, 512 Hz and 1024 Hz. **(B)** Time domain grand averaged EFRs for each modulation frequency collected for all participants from this study. The neural response 'follows' the stimulus amplitude envelope shown in A, most clearly seen for the slower modulation frequencies. **(C)** Responses in (B) transformed to the frequency domain using a fast-Fourier transform exhibits clear peaks at all modulation frequencies tested in this study. **(D)** Signal to noise ratio of EFR peaks in the frequency domain for all participants and all modulation frequencies suggests an average SNR of 6 dB across all modulation frequencies.

EFR amplitudes did not differ between Fz-R and Fz-L montages ($\beta = -0.003$, $t = -1.318$, $p = .198$). Trends seen at 512 Hz were amplified for 1024 Hz AM, where EFR amplitudes in the Fz-R montage were greater than Fz-L montage ($\beta = -0.004$, $t = -5.139$, $p < .001$) but did not differ from amplitudes in the L-R montage ($\beta < 0.001$, $t = 0.285$, $p = .777$), suggesting more distal neural generators in the peripheral auditory pathway.

Taken together, these results suggest that EFRs to varying AM rates emphasize peripheral versus central neural generators. Further, the Fz-R montage was best suited among our montages to capture EFRs along the entire auditory neuraxis for all AM rates.

Young listeners with normal audiometric thresholds exhibited substantial variability in both EFR amplitudes and speech perception in noise

Participants demonstrated individual differences in EFR amplitudes in the Fz-R montage and the shape of the temporal modulation transfer function (i.e., EFR amplitudes as a function of AM rate; Fig. 3A; Table 1). To obtain a single metric of the temporal modulation transfer function per participant, we used a growth curve analysis (GCA[39]) to calculate an EFR slope across all four AM rates (Fig. 3B). Growth curve analysis uses orthogonal polynomial time terms to model change over time, which can be used to estimate the temporal modulation transfer function from peripheral to central generators. EFR amplitudes follow a low-pass shape as a function of increasing modulation frequency[23,40]. Hence, we expect EFR slopes to be negative in direction (Fig. 3A). Steeper slopes (i.e., more negative), would indicate greater change from peripheral to central generators, while flatter slopes would suggest lesser relative increases in amplitudes from peripheral to central generators. We then used individual participant EFR slopes to compare the EFRs to other measures across this study. Mean centered EFR slopes ranged from −0.142 to 0.389. These data suggested that there was substantial variability in EFR amplitudes across AM rates, even in young adults with normal audiometric thresholds. Presumably, this variability in EFR amplitudes reflects individual differences in the encoding of temporal information along the ascending auditory pathway.

We then examined speech perception in noise using QuickSIN, a clinically used multi-talker speech intelligibility task with moderate contextual and linguistic cues (Fig. 3C). Each participant's speech perception in noise score was calculated as the proportion of correct keywords identified per SNR level across all four test lists.

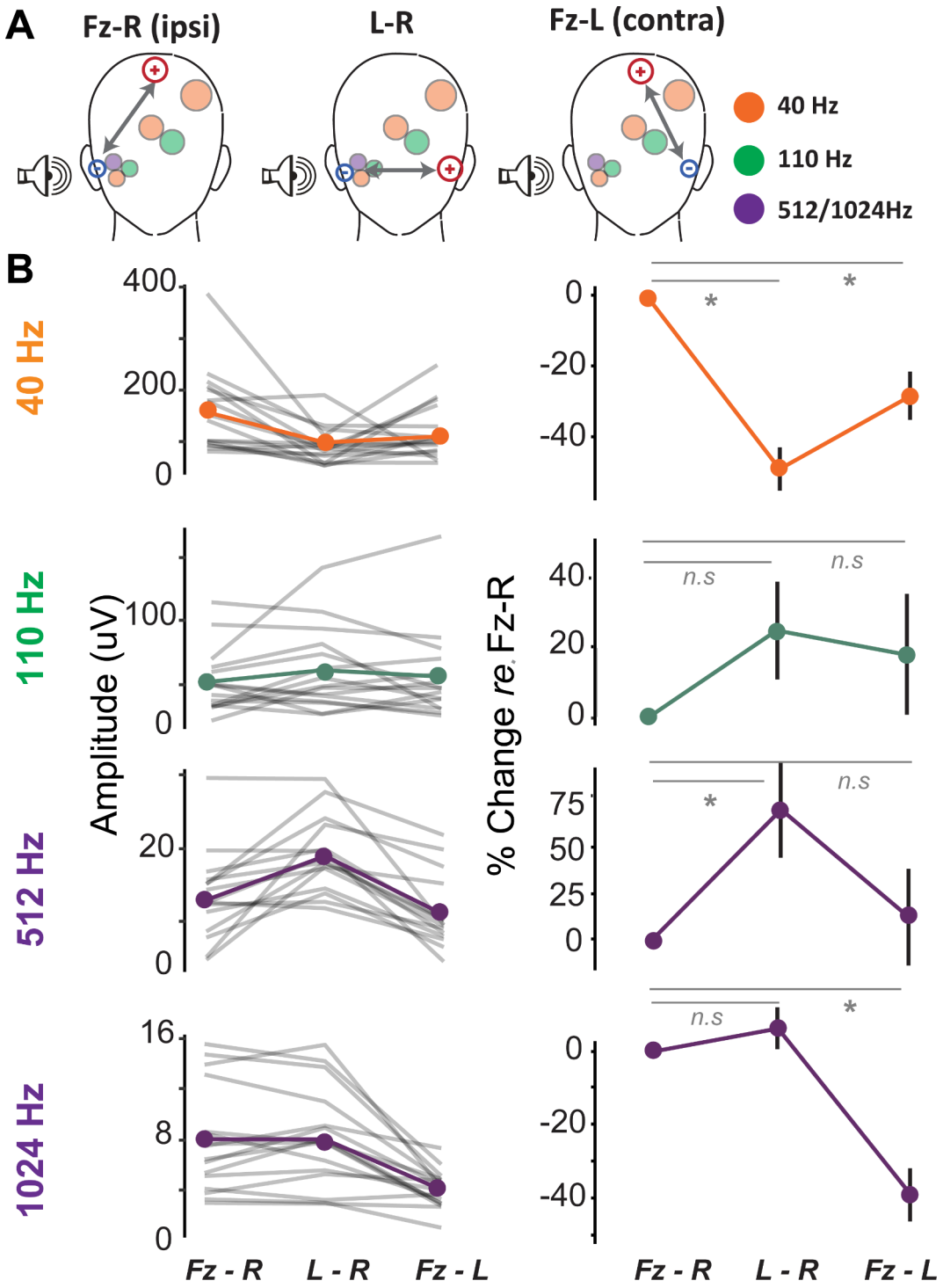

Fig. 2. Electrode montage configurations emphasize complementary neural generators. **(A)** Schematic of active and reference electrode placement of three montage configurations: Fz to Right-tiptrode (Fz-R), Left-tiptrode to Right-tiptrode (L-R), and Fz to Left-tiptrode (Fz-L). The shaded circles represent the auditory nuclei that generate these responses, with circle size reflecting the size of the generators. **(B)** Left: Average envelope following responses (EFR) amplitudes for each AM rate and each montage configuration. Right: The percent change in EFR amplitude at each AM rate for L-R and Fz-L configurations, relative to Fz-R.

Clinical scores ranged from −4 to 1.25 dB SNR loss, which were all within clinically normal limits[34]. However, behavioral performance at SNR 0 dB was significantly lower than performance at the easier SNRs of 20, 15, 10, and 5 dB ($ps < 0.001$, Linear mixed-effects model, Fig. 3C; Table 2). On average, participants had near-perfect performance for SNRs at 20, 15, 10 and 5dB, ranging from an average of 99% at SNR 20 to an average of 94% at SNR 5. Performance at these SNRs between 20 and 5 dB did not significantly differ from one another ($ps > 0.05$,

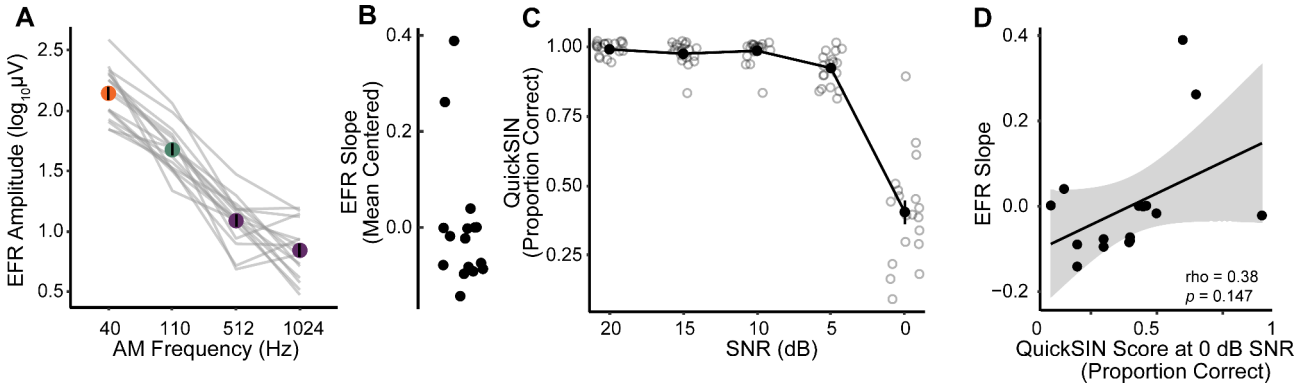

Fig. 3. Young listeners with normal hearing thresholds exhibit substantial individual differences in EFRs and speech perception in noise. **(A)** Temporal modulation transfer function with EFR amplitudes at each modulation rate. Individual participant lines are denoted in gray. **(B)** Distribution of individual EFR slope estimates for each participant. **(C)** Proportion of keywords identified correctly across four QuickSIN lists for each SNR level. Solid line and points denote average performance across all participants. Smaller points denote individual participant scores. QuickSIN performance significantly drops from SNR 5 to SNR 0 dB. **(D)** The scatterplot reveals a non-significant correlation between EFR slope estimates and QuickSIN scores at signal-to-noise ratio (SNR) 0 dB.

Fixed Effects	Estimate	SE	t-value	*p*-value
Intercept	−1.604	0.016	−99.00	< 0.001 ***
ot1	−0.966	0.032	−30.12	< 0.001 ***

Table 1. Fixed effect estimates for model of EFR amplitudes across amplitude modulation rates (observations = 62). ***$p < .001$. Growth curve model: *lmer(EFR Amplitude (log) ~ ot1 + (0 + ot1|participant), control = lmerControl(optimizer = 'bobyqa'), REML = FALSE)*.

Contrast	Estimate	SE	t-value	*P*
SNR 0 vs. SNR 5	−0.534	0.033	−16.198	< 0.001 ***
SNR 0 vs. SNR 10	−0.581	0.033	−17.618	< 0.001 ***
SNR 0 vs. SNR 15	−0.578	0.033	−17.524	< 0.001 ***
SNR 0 vs. SNR 20	−0.588	0.033	−17.808	< 0.001 ***
SNR 5 vs. SNR 10	−0.047	0.033	−1.421	0.268
SNR 5 vs. SNR 15	−0.044	0.033	−1.326	0.271
SNR 5 vs. SNR 20	−0.053	0.033	−1.610	0.225
SNR 10 vs. SNR 15	0.003	0.033	0.095	0.925
SNR 10 vs. SNR 20	−0.006	0.033	−0.189	0.925
SNR 15 vs. SNR 20	−0.009	0.033	−0.284	0.925

Table 2. Multiple comparisons from a linear mixed effects model examining QuickSIN performance. *p*-values were adjusted using the Benjamini-Hochberg Procedure.

Table 2). Performance dropped significantly for most participants at the most challenging SNR of 0 dB, and this drop in performance was statistically significant (ps < 0.001, Table 2). Performance at the most challenging SNR also varied widely, with percent correct ranging from 10 to 90% in our participants ($M = 40.3\%$, $SD = 20.4\%$). These results suggest that while young listeners performed near ceiling at easier listening conditions, there was significant individual variability under challenging listening conditions, despite normal audiometric thresholds.

Finally, we explored if there were any correlations between performance on QuickSIN and neural encoding metrics obtained from the EFRs. QuickSIN performance at 0 dB SNR did not significantly correlate with EFR amplitudes at 40 Hz ($rho = 0.180$, $p = .496$), 110 Hz ($rho = -0.067$, $p = .813$), 512 Hz ($rho = 0.120$, $p = .678$), or 1024 Hz ($rho = 0.240$, $p = .380$). Further, QuickSIN performance at 0 dB SNR did not significantly correlate with the EFR slope metric (Fig. 3D). These results suggest that there was no direct, linear relationship between the EFR bottom-up measures of sensory encoding and QuickSIN performance within our participant population.

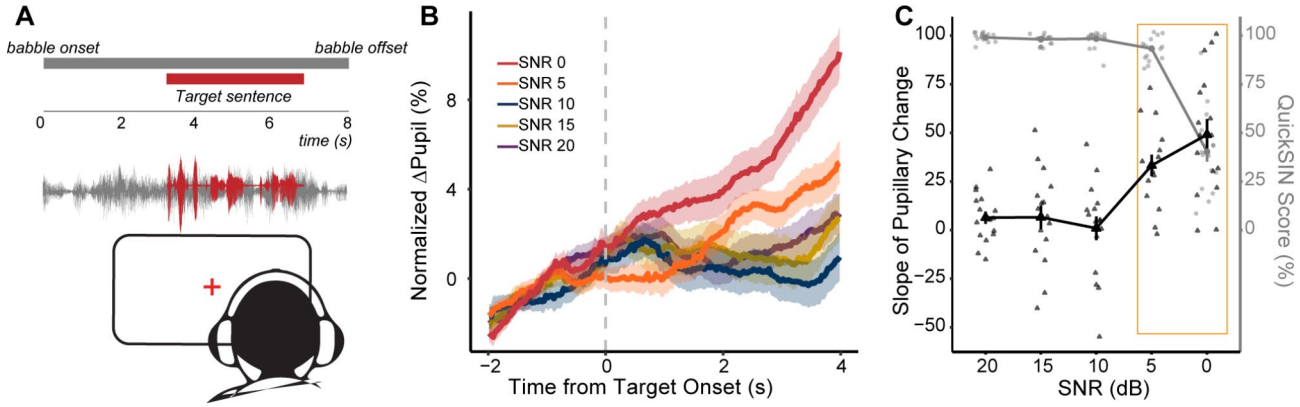

Fig. 4. Pupillary responses to QuickSIN show significant changes in listening effort with increasing SNR difficulty. **(A)** Schematic of experimental design. Subjects watched a fixation point and had changes in their pupil diameters measured while listening to QuickSIN sentences. **(B)** Grand-averaged raw pupillary responses to QuickSIN at each SNR level time-locked to the onset of the target speaker. **(C)** Average QuickSIN performance (right y-axis) at each SNR level in gray, overlaid with the average pupillary slope (left y-axis) at each SNR level in black. Individual participant QuickSIN scores and pupillary slopes are denoted by lighter circles and triangles, respectively. Listening effort, as measured by the pupillary slopes, at SNR 5 dB significantly increased while QuickSIN performance remained near ceiling, yet the increase in listening effort at SNR 0 dB was not associated with better performance at SNR 0 dB (orange square).

Contrasts	Estimate	SE	t-value	*p*-value
SNR 0 vs. SNR 5	3.344	1.498	2.232	0.064
SNR 0 vs. SNR 10	5.853	1.474	3.971	0.001 **
SNR 0 vs. SNR 15	4.643	1.474	3.150	0.006 **
SNR 0 vs. SNR 20	4.590	1.474	3.114	0.006 **
SNR 5 vs. SNR 10	2.509	1.498	1.674	0.188
SNR 5 vs. SNR 15	1.298	1.498	0.867	0.457
SNR 5 vs. SNR 20	1.246	1.498	0.831	0.457
SNR 10 vs. SNR 15	−1.210	1.474	−0.821	0.457
SNR 10 vs. SNR 20	−1.263	1.474	−0.857	0.457
SNR 15 vs. SNR 20	−0.053	1.474	−0.036	0.971
ot1 x SNR 0 vs. SNR 5	0.026	0.014	1.905	0.081
ot1 x SNR 0 vs. SNR 10	0.080	0.014	5.878	< 0.001 ***
ot1 x SNR 0 vs. SNR 15	0.071	0.014	5.190	< 0.001 ***
ot1 x SNR 0 vs. SNR 20	0.071	0.014	5.211	< 0.001 ***
ot1 x SNR 5 vs. SNR 10	0.054	0.014	3.877	< 0.001***
ot1 x SNR 5 vs. SNR 15	0.044	0.014	3.200	0.002 **
ot1 x SNR 5 vs. SNR 20	0.045	0.014	3.221	0.002 **
ot1 x SNR 10 vs. SNR 15	−0.009	0.014	−0.688	0.561
ot1 x SNR 10 vs. SNR 20	−0.009	0.014	−0.667	0.561
ot1 x SNR 15 vs. SNR 20	< 0.001	0.014	0.021	0.983

Table 3. Multiple comparisons contrasts for the GCA for overall change in the pupillary responses between SNR levels and the slopes of the pupillary responses. $^{**}p < .01$; $^{***}p < .001$. *p*-values were adjusted using the Benjamini-Hochberg procedure.

Increases in listening effort are evident prior to changes in behavioral thresholds

Given the lack of a correlation between the EFRs and performance at SNR 0 on QuickSIN, we then asked whether top-down measures of listening effort changed with SNR and if those changes were more indicative of QuickSIN performance. Isoluminous task related changes in pupil diameters were measured as participants completed the task at various SNR levels (Fig. 4A). Pupillary responses increased after babble onset and during the presentation of the stimuli. We used a GCA to calculate the change in the pupillary response over time during listening, time-locked to target speech onset (i.e., 3 s after babble onset). Pupil sizes were modulated by SNR level, with SNR 0 and SNR 5 showing the fastest change in pupil size in our study population (Table 3; Fig. 4B). This suggests that these SNR levels required the greatest amount of listening effort compared to the others in our study.

Individual participant's pupillary slopes at each SNR were then extracted from the random effect of the GCA. Overlaying the average pupillary slopes at each SNR against task performance, we found that pupillary slopes were relatively flat for SNRs 20, 15 and 10 dB where task performance was near ceiling (Fig. 4C). Interestingly, pupillary slopes were significantly greater at SNR 5 dB relative to SNR 10, SNR 15, and SNR 20 (Table 3), even though performance was still near ceiling (Table 2). There was another increase in the steepness of the slope at SNR 0, where QuickSIN scores also dropped significantly, although the difference in the slopes between SNR 5 and SNR 0 was not statistically significant after correcting for multiple comparisons ($p = .081$). While these data suggest that an increase in task difficulty co-occurred with an increase in listening effort as indexed by pupillometry, it is interesting to note that pupillometric changes were not necessarily reflective of behavioral performance outcomes. The increase in the pupillary slope at SNR 5 suggests an increase in listening effort that was enough to maintain near-ceiling performance, but the increase in listening effort at SNR 0 did not counteract the detrimental effects of noise, resulting in a decreased performance relative to easier SNRs.

Sensory coding measures and listening effort provide complementary contributions to the variance in speech perception in noise

Given the results seen above, we focused further analyses on the most challenging two SNRs – (1) SNR 0 where effort increased but performance dropped, and (2) SNR 5 where performance was near-ceiling, but effort increased. Individual participant pupillary slopes at SNR 0 were not significantly correlated with task performance at 0 dB SNR (Fig. 5A) or 5dB SNR ($rho = -0.030$, $p = .917$). Additionally, these pupillary responses at SNR 0 were not significantly correlated with individual EFR amplitudes (40 Hz: $rho = 0.07$, $p = .795$; 110 Hz: $rho = -0.120$, $p = .666$; 512 Hz: $rho = -0.120$, $p = .657$; 1024 Hz: $rho = -0.380$, $p = .144$;) or EFR slope (Fig. 5B). These data suggest that there was no direct relationship between the neural encoding measures, listening effort, and multi-talker speech intelligibility within our participant population.

To probe the potential synergistic effect between sensory coding and listening effort, we used a forward and backward stepwise regression to calculate the variance explained by EFRs and pupil diameters (i.e., effort) on performance in the QuickSIN task. For multi-talker speech performance at SNR 0, the stepwise linear regression revealed that EFR slope alone ($SpIN\ SNR0 \sim EFR$) provided the best model fit, with an Adjusted-R^2 = 0.133 and Akaike information criteria (AIC) = −3.922. This suggests that EFR slopes explained approximately 13.25% of the variance in QuickSIN performance at SNR 0. However, this best-fit model was not statistically significant ($F_{(1,14)} = 3.292$, $p = .091$). The regression steps that included only pupillary slope at SNR 0 or the combination of pupillary slopes at SNR 0 and EFR slope explained 1.8% (AIC = −1.936) and 14.3% (AIC = −3.301) of the variance in QuickSIN at SNR 0, respectively and were also not statistically significant models (Fig. 5C). These results suggest that the pupillary slopes at SNR 0 did not contribute significantly to the variance in task performance at that SNR.

We then asked whether effort at SNR 5, when performance was still near ceiling, was a better predictor of performance at SNR 0. When we constructed the stepwise regression model to explain performance at 0 dB SNR with just the pupillary slope at SNR 5 instead of the pupillary slope at SNR 0, it revealed that the combination of pupillary slope at SNR 5 and EFR slopes were the best fit model (AIC = −49.333). This best-fit model was statistically significant ($F_{(2,12)} = 3.92$, $p = .049$) with a moderate effect size ($f^2 = 0.653$) and 69.2% power. The intermediary step with just pupillary slope at SNR 5 explained 18.7% of the variance in QuickSIN at SNR 0 (AIC = −3.437). The addition of EFR slopes to this model increased the Adjusted-R^2 to 29.4% and an AIC of −4.765, yielding the best fit model. Hence, these results suggest that sensory neural encoding and listening effort

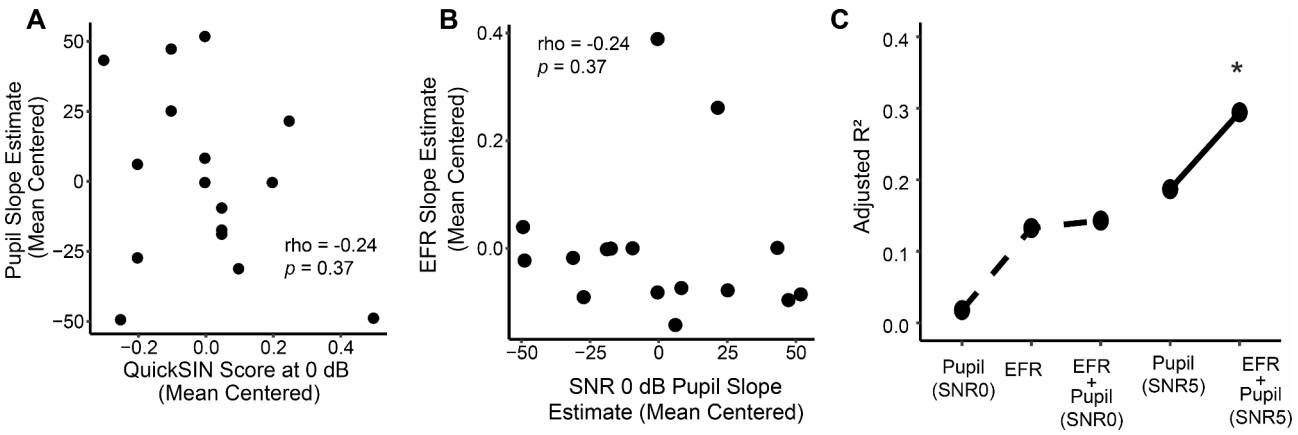

Fig. 5. Pupillary slopes and the slope of envelope following responses (EFR) significantly contribute to multi-talker speech intelligibility. **(A)** Pupillary slopes at SNR 0 dB (mean centered) are not associated with QuickSIN performance at SNR 0 dB. **(B)** Scatterplot between envelope following response (EFR) slopes and pupil slopes at SNR 0 dB. **(C)**A stepwise regression measuring the variance explained (Adjusted-R^2) of QuickSIN at SNR 0 dB revealed incremental improvement with the combination of both EFR slope and pupil slope at SNR 0 dB (dashed line). A second stepwise regression showed an even greater increase in the variance explained in QuickSIN at SNR 0 dB when including the pupil slope at SNR 5 dB (solid line).

provided complementary contributions to the overall variance explained in speech perception in noise. Further, the amount of listening effort required to maintain performance at the relatively easier SNR 5 is a better indicator of an individual's multi-talker speech intelligibility at the more challenging SNR 0.

Discussion

Speech intelligibility is an inherently multi-dimensional process that includes exquisite interactions between sensory-motor systems, language networks, and top-down attentional networks that include cognition and arousal[41–51]. A combination of influences due to diverse factors such as peripheral hearing thresholds, cochlear health, central auditory processing, arousal state, cognitive status, current emotional status, and familiarity with context of the conversation affects multi-talker speech intelligibility[48,52–58]. Studies that probe the individual effects of these factors by carefully controlling for as many of these factors as possible can only explain a small part of the overall individual variance in speech perception performance. Here, we explored the synergistic contributions of two factors affecting multi-talker speech intelligibility – bottom-up sensory coding fidelity to temporally modulated sounds that was indexed using EFRs and top-down cognitive load that was indexed using pupillometry. We found that although these two factors individually contributed to only a small percentage of the overall variance in our multi-talker speech task, they combined synergistically to explain a significant portion of the variance in performance on our task (Fig. 5C). These results are a first step towards a multi-dimensional approach to understanding speech perception in noise.

EFRs have been used in multiple studies to assess neural coding to the temporally modulated stimulus amplitude envelope[40,51,59–63]. The neural generators of EFRs evoked to speech sounds, also referred to more generically as the frequency-following response, have been a subject of recent debate. Historically considered to have primarily subcortical generators, recent evidence suggests a larger cortical contribution to these responses[64–68]. In contrast, EFRs evoked to simpler sinusoidally amplitude modulated tones have long been known to emphasize cortical or subcortical generators depending on the AM frequency used, with slower (< 40 Hz) AM rates thought to emphasize cortical generators, and faster (100–300 Hz) AM rates thought to emphasize subcortical generators[18,21,69]. Recent studies also suggest much faster AM rates to be sensitive to peripheral neural degeneration, potentially stemming from the auditory nerve[8,19,70]. These differential neural generators are possibly due to a fundamental biophysical property of the neurons within the auditory pathway, wherein the phase-locking abilities of the auditory nerve neurons extend up to approximately 2000 Hz, but this upper limit gradually decreases along the afferent pathway to about 80 Hz in the auditory cortex[20].

While human studies only use a few modulation rates due to limitations of recording time, animal studies have systematically characterized these EFR generators using lesioning, anesthesia or peripheral nerve damage[8,19,23,69]. Human studies also typically do not use modulation rates faster than ~ 300 Hz, as the decrease in EFR amplitudes with rate is logarithmic, reaching very small amplitudes at faster rates. Here, through a combination of high sampling rate, ear-canal electrodes, and a low recording noise floor, we demonstrated that we can reliably record EFRs to AM rates up to 1024 Hz in young listeners with normal hearing thresholds (Fig. 1). We further compared EFRs across multiple electrode montages to support the idea that faster AM rates emphasize more peripheral generators (Fig. 2). Our findings are in agreement with previous human and animal model studies that used similar multi-channel approaches to determine EFR generators[23,35,36,71–73]. We further used a GCA to determine the slope of EFR change with modulation frequency, providing a metric for assessing temporal processing along the ascending auditory pathway (Fig. 3A-C). This slope metric also has the added benefit of minimizing inter-subject variability due to recording factors such as head size or electrode impedance, as they are normalized within subject by design. There was a trend for steeper EFR slopes to be associated with poorer QuickSIN performance at 0 dB SNR (Fig. 3D), though this was not statistically significant. Because EFR slopes may be affected by both peripheral neural coding and potential compensatory central gain, future studies will explore the roles of these individual contributors to speech in noise intelligibility in a larger sample.

Individuals showed significant variability in performance on QuickSIN, particularly at the most challenging SNR of 0dB, despite being young and having clinically normal hearing thresholds (Fig. 3D). QuickSIN is clinically used to primarily assess speech perception in noise abilities in listeners with hearing loss, using open-set sentences with linguistic context masked in multi-talker babble[34]. However, adults with normal hearing can still show a large variability in QuickSIN performance[74]. Clinical assessment of QuickSIN assesses SNR loss, or the SNR level at which listeners can achieve 50% accuracy[34], which may mask performance variability at specific SNR levels in adults with normal hearing. Here, we specifically examined QuickSIN performance at individual SNR levels to examine how listening effort changed with increasing listening difficulty. Our QuickSIN scores are consistent with prior research showing similar variability in multi-talker speech intelligibility in adults with normal hearing using a digits task consisting of a target speaker and two competing co-localized speakers producing speech streams of a closed set of monosyllabic numbers devoid of linguistic context[2]. Language experience can also impact speech in noise intelligibility[75], but all participants in this study were self-reported fluent speakers of English. Additional information on other language experience was not collected, yet if language experience was a significant factor in QuickSIN performance, we would expect to see greater variability at other SNR levels besides 0 dB and QuickSIN dB SNR losses outside the normal range.

Listening effort, assessed using pupillometry, increased with more challenging SNRs on QuickSIN, with SNR 0 and SNR 5 resulting in the greatest change in pupil diameter. Pupillometry has rapidly gained prominence as a tool to assess cognitive load and effortful listening, which is modulated by task difficulty, cognitive status and hearing acuity[31,48,76–80]. The precise neural pathways that induce pupillary changes during listening is still under study, but it is hypothesized to be mediated by the locus coeruleus–norepinephrine (LC-NE) system. The LC-NE system is a network of neurons that has wide-spread projections throughout the cortex. Changes in LC-NE system activity are strongly associated with changes in pupil diameter[81]. However, the underlying neural mechanisms modulating pupillary changes may encompass networks that go beyond the LC-NE system, and

may be driven by other networks that modulate arousal states[32,82]. Animal studies demonstrate some support for this idea, with pupil diameter indexing momentary changes in arousal states and indicative of task performance under challenging listening conditions[30,81,83]. These studies suggest a non-linear relationship between pupil-indexed arousal and behavior.

Optimal states of arousal result in improved behavioral outcomes. However, task-evoked pupillary changes that are lower than the optimal state result in disengagement, and changes that are higher than the optimal state result in hyperarousal, both of which affect behavioral performance[30]. This potential tradeoff between effort and performance may follow an inverted U-shaped curve[26]such that increases in effort can benefit performance, but only to a certain extent. That is, when a listening situation becomes too difficult, increased listening effort is not necessarily beneficial[26]. The non-monotonic relationship between task-evoked pupil size and listening effort[84]may also help to explain our findings, wherein pupillary slopes at SNR 0 were not predictive of performance at this SNR level. Rather, task-evoked pupil diameter at SNR 5dB explained more variance in behavioral performance at SNR 0. This finding had a moderate effect size ($f^2 = 0.653$, ~70% power) and was statistically significant, even in our limited sample size of sixteen participants. This suggests that an individual who had reached their optimal, intermediate listening effort level at SNR 5 would have likely experienced a decrease in performance at SNR 0. Speech perception at SNR 0 for these individuals was likely too difficult, such that an increase in listening effort at SNR 0 did not benefit performance. Conversely, an individual who was approaching the optimal, intermediate listening effort level at SNR 5 would likely experience less of a decrease in performance at SNR 0 compared to someone who already surpassed their optimal performance level by SNR 5. A follow up study with a larger cohort of younger and middle-aged adults with normal audiograms further replicate these relationships between listening effort and behavioral performance[85]. Thus, our results suggest that optimal listening effort to benefit speech in noise intelligibility occurs at moderately difficult SNR levels and this effort can be predictive of performance at more difficult SNR levels.

Increasing evidence suggests that extended high frequency hearing in listeners with normal hearing at lower frequencies contribute to speech perception in noise abilities[86–88]. In an exploratory analysis, we examined the potential contributions of subclinical variability in pure tone averages at 0.5, 1, 2, and 4 kHz and extended high frequency pure tone averages at 12 and 16 kHz to the variability observed in QuickSIN performance, listening effort, and EFR amplitudes (see Supplementary Material 1). Pure tone averages in the clinical and extended high frequency ranges were not significant covariates in any of the analyses, suggesting that the variability in QuickSIN, listening effort, and EFR amplitudes could not be explained by individual differences in hearing sensitivity in our young adults with clinically normal hearing.

In our study, neither the EFR slopes, nor the pupillary changes were directly related to QuickSIN task performance. Yet, together, they provided a greater proportion of variance explained compared to either metric in isolation. Pupillary dilations increased significantly for SNR 5, even though there was no added benefit or detriment seen in behavioral performance at that SNR (Fig. 4B). Interestingly, this change in pupil diameter at SNR 5 was a greater indicator for behavioral performance at SNR 0, compared to pupillary changes at SNR 0 itself (Fig. 5C). This suggests that perhaps the ability to expend effort to maintain performance at a less challenging SNR is more predictive of performance at harder SNRs. Future work will explore these interactions between SNRs and pupillary changes, as well as assess if either the sensory coding component or the top-down indices of effortful listening change independently or concurrently with hearing pathologies.

Methods

Participants

Nineteen English-speaking participants (mean age = 28.70, *SD* = 4.15 years) were recruited from the greater Boston, MA area to complete all portions of the study. Participants were self-reported proficient in English. Testing occurred over two sessions separated by less than one week apart. Three participants did not complete the electrophysiological session of the experiment, resulting in a final sample size of sixteen (6 male, 10 female). Cognition[89,90], depressive symptoms[55,91,92], and perhaps tinnitus[93–95]have each been shown to impact speech perception. Therefore, participant eligibility was determined on the first visit by screening for cognitive skills (Montreal Cognitive Assessment > 25[96]), depression (Beck's Depression Inventory < 21[97]), and tinnitus (Tinnitus reaction questionnaire < 72[98]). All participants had normal hearing sensitivity with air conduction thresholds ≤ 25dB for octave frequencies between 250 and 8000 Hz and were not users of assistive listening devices ("Do you routinely use any of the following devices – cochlear implants, hearing aids, bone-anchored hearing aids or FM assistive listening devices?", participant has to answer No for inclusion). Participants received monetary compensation per hour for their participation. This research protocol was approved by the Institutional Review Board at Massachusetts Eye and Ear Infirmary (Protocol #1006581) and Partners Healthcare (Protocol #2019P002423). All procedures were performed in accordance with the relevant guidelines and regulations therein. All participants provided informed consent.

Pupillometry

Stimuli and acquisition

Pupillary responses were recorded with a head mounted pupillometry system at a 30 Hz sampling rate (Argus Science ET-Mobile) while participants completed the Quick Speech in Noise (QuickSIN) test[34]. QuickSIN is a standard test of speech perception in noise that provides a measure of signal-to-noise ratio (SNR) loss, which indicates the lowest SNR level at which the listener can accurately identify words 50% of the time. The tests were administered using a Windows Surface tablet at a fixed distance from the participant. Each QuickSIN test list consisted of six sentences presented monoaurally to the right ear at 70 dB SPL masked in four-talker babble. The intensity of the four-talker babble was modulated to produce the following SNR levels: 25, 20, 15, 10, 5, and 0 dB.

Sentences were presented in descending order based on SNR level to match the testing procedure for QuickSIN in audiology clinics[34]. All participants completed two practice QuickSIN lists before completing four test lists. Participants were instructed to fixate on a point on the screen during listening and to repeat the target sentence to the best of their ability. Each target sentence contained five keywords for identification. The number of key words identified per sentence were recorded. Then, the proportion of keywords correctly identified for each SNR across all four test lists (20 total key words per SNR) was calculated for each participant.

Prior to QuickSIN testing, the dynamic range of the pupil was first characterized in each participant by sinusoidally varying the grayscale intensity of the screen from 0 (black) to 255 (white) at 0.05 Hz (i.e., 20 s periods). Four cycles of this dynamic range were presented. The screen brightness was then set to midpoint gray, and the ambient lighting in the testing room was adjusted to obtain a baseline pupil size in the middle of the dynamic range for each participant.

Pupillometry processing and analysis

Pupillary responses time-locked to the onset of the QuickSIN sentences were processed as described by Winn et al.[99]. Blinks and saccades were linearly interpolated from approximately 120 ms before to 120 ms after the detected noise[33,100]. Any trial containing blinks or saccades that were longer than 600 ms were removed from further analysis[101]. Baseline pupil size was calculated as the mean pupil size in the 3s prior to target speaker onset. Each data point was then normalized on a trial-by-trial basis by first subtracting the baseline pupil size then dividing by the baseline. The resultant pupillary data is expressed as the percent change in pupil size relative to baseline.

The pupillary responses were then averaged across all four test lists at each SNR level for each participant. We excluded SNR 25 from all further analyses, as visual inspection of the averaged pupillary responses at SNR 25 showed different temporal dynamics compared to pupillary responses for the other SNRs. The difference in temporal dynamics at SNR 25 was likely due to the time related to familiarization for the task in each block, as SNR 25 was always presented first in each list. Pupillary responses time-locked to the onset of the target speaker were analyzed for the remaining SNR levels.

A growth curve analysis (GCA[39]; was used to obtain a measure of the slope of the pupillary response during listening from the onset of the target speaker. GCA uses orthogonal polynomial time terms to model distinct functional forms of the pupillary response over time. A GCA was fit using a first-order orthogonal polynomial to model the interaction with SNR level. This first-order model provided two parameters to explain the pupillary response. The first is the intercept, which refers to the overall change in the pupillary response over the time-window of interest. The second is the linear term (ot1), which represents the slope of the pupillary response over time. The GCA model included fixed effects of SNR (20, 15, 10, 5, and 0; reference = 0) on the linear term with a random slope of the interaction between participant and SNR on the linear time term:

$$Pupil \sim ot1{*}SNR + (ot1 \mid Subject : SNR)$$

GCA were conducted in R[102] using the *lme4*package[103], and *p*-values were estimated using the *lmerTest*package[104]. Multiple comparisons were performed using the *emmeans* package. Adjusted *p*-values are reported using Benjamini-Hochberg Procedure to control for the false discovery rate[105].

Electrophysiology

Stimuli and acquisition

Electroencephalography (EEG) for the EFRs was collected in an electrically shielded sound attenuating chamber. Participants were seated in a reclined chair and were instructed to minimize movements. The recording session lasted approximately three hours and participants were given breaks as necessary. Recordings were collected using a 16-channel EEG system (Processor: RZ6, preamplifier: RA16PA, differential low impedance amplifier: RA16-LID, TDT System) with two gold-foil tiptrodes positioned in the ear canals (Etymotic) with a sampling rate of 24,414.0625 Hz. Cup electrodes were placed at the Fz site and at both ear lobes, all referenced to a ground at the nape. Electrode impedances were below 1 kΩ after prepping the skin (NuPrep, Weaver and Co.) and applying a conductive gel between the electrode and the skin (EC2, Natus medical).

Envelope following responses (EFRs) were recorded to amplitude modulated (AM) tones. The stimulus carrier frequency was 3000 Hz with amplitude modulation rates of 40, 110, 512, and 1024 Hz. The 3000 Hz carrier frequency was chosen to allow the use of modulation frequencies up to 1024 Hz, while remaining within frequencies considered to be relevant to everyday communication. Stimuli were 200 ms long, presented with a 3.1 per second repetition rate in alternating positive and negative polarities. A calibrated ER3A insert earphone in the right ear was used for stimulus presentation. Stimuli presentation level was 85 dB SPL. Stimulus delivery (sampling rate: ~100 kHz) and signal acquisition were coordinated using the TDT system low impedance amplifier and a presentation and acquisition software (LabVIEW).

Electrophysiology processing and analysis

EFRs were processed using a fourth-order Butterworth filter with a lowpass filter of 3000 Hz. The following highpass filter cutoffs were used for 40 Hz, 110 Hz, and 1024 Hz AM stimuli, respectively: 5 Hz, 80 Hz, and 300 Hz. Fast Fourier transforms (FFTs) were performed on the averaged time domain waveforms for each participant at each AM rate starting 10 ms after stimulus onset to exclude ABRs and ending 10ms after stimulus offset using MATLAB v2022a (MathWorks Inc., Natick, Massachusetts). The maximum amplitude of the FFT peak at one of three adjacent bins (~3 Hz) around the modulation frequency of the AM rate is reported as the EFR amplitude. FFT amplitudes at ten frequency bins on either side of the peak 3 bins were averaged to calculate

the noise floor. The signal to noise ratio was calculated as the ratio of the max FFT amplitude at one of the three center bins to the noise floor, expressed in dB.

A GCA was then fit to model EFR amplitudes across each AM rate. The best-fit model that promoted model convergence and did not produce a singular fit contained a first-order orthogonal polynomial with a random slope of the linear time term (ot1) per participant that removed the correlation between the random effects:

$$EFR \sim ot1 + (0 + ot1 \mid Subject)$$

Statistical analysis

Correlations between QuickSIN scores with EFR metrics and pupillary metrics were assessed using Spearman's rank correlations due to non-normal distributions in the data. To examine speech intelligibility scores, a linear mixed effects model was fit with QuickSIN score as the outcome variable, a fixed effect of SNR level, and a random effect of participant. Multiple comparisons between SNR levels were performed using the *emmeans*package in R[102]. Adjusted *p*-values are reported using the Benjamini-Hochberg Procedure to control for the false discovery rate[105].

A stepwise multivariate regression was performed to explain variability in QuickSIN at SNR 0 due to EFRs and pupillary responses. The outcome variable was QuickSIN performance at SNR 0, with the pupillary slope at SNR 0 and the EFR slope as predictor variables. The stepwise regression selected the best-fit model based on AIC and adjusted-R^2 for each model that was calculated. The linear regression was performed in R using *lme4*package[103]. The stepwise regression was performed using the *MASS*package[106]. A post-hoc power analysis of the regression was performed using GPOWER[107].

Data availability

The datasets generated during and/or analyzed during the current study are available in the Open Science Framework repository and can be accessed at https://osf.io/nt7ep/.

Received: 4 March 2024; Accepted: 28 November 2024
Published online: 28 December 2024

References

1. Hind, S. E. et al. Prevalence of clinical referrals having hearing thresholds within normal limits. *Int. J. Audiol.* **50**, 708–716 (2011).
2. Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V. & Polley, D. B. Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing. *eLife* **9**, e51419 (2020).
3. Spehar, B. P. & Lichtenhan, J. T. Patients with normal hearing thresholds but Difficulty hearing in noisy environments: a study on the willingness to try auditory training. *Otol Neurotol.* **39**, 950–956 (2018).
4. Kujawa, S. G. & Liberman, M. C. Adding insult to Injury: cochlear nerve degeneration after 'Temporary' noise-Induced hearing loss. *J. Neurosci.* **29**, 14077–14085 (2009).
5. Kujawa, S. G. & Liberman, M. C. Synaptopathy in the noise-exposed and aging cochlea: primary neural degeneration in acquired sensorineural hearing loss. *Hear. Res.* (2015). (ePub ahead of print).
6. Lobarinas, E., Salvi, R. & Ding, D. L. Insensitivity of the audiogram to carboplatin induced inner hair cell loss in chinchillas. *Hear. Res.* **302**, 113–120 (2013).
7. Sergeyenko, Y., Lall, K., Liberman, M. C. & Kujawa, S. G. Age-related cochlear synaptopathy: an early-onset contributor to Auditory Functional decline. *J. Neurosci.* **33**, 13686–13694 (2013).
8. Parthasarathy, A. & Kujawa, S. G. Synaptopathy in the aging cochlea: characterizing early-neural deficits in auditory temporal envelope processing. *J. Neurosci.* https://doi.org/10.1523/jneurosci.3240-17.2018 (2018).
9. Wu, P. Z. et al. Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. *Neuroscience* https://doi.org/10.1016/j.neuroscience.2018.07.053 (2018).
10. Chambers, A. R. et al. Central Gain restores auditory Processing following Near-Complete Cochlear Denervation. *Neuron* **89**, 867–879 (2016).
11. Auerbach, B. D., Radziwon, K. & Salvi, R. Testing the Central Gain Model: loudness growth correlates with Central Auditory Gain Enhancement in a Rodent Model of Hyperacusis. *Neuroscience* **407**, 93–107 (2019).
12. Parthasarathy, A., Bartlett, E. L. & Kujawa, S. G. Age-related changes in neural coding of envelope cues: peripheral declines and central compensation. *Neuroscience* **407**, 21–31 (2019).
13. Parthasarathy, A., Herrmann, B. & Bartlett, E. L. Aging alters envelope representations of speech-like sounds in the inferior colliculus. *Neurobiol. Aging.* **73**, 30–40 (2019).
14. Resnik, J. & Polley, D. B. Cochlear neural degeneration disrupts hearing in background noise by increasing auditory cortex internal noise. *Neuron* https://doi.org/10.1016/j.neuron.2021.01.015 (2021).
15. McGill, M. et al. Neural signatures of auditory hypersensitivity following acoustic trauma. *Elife* **11**, e80015 (2022).
16. Rumschlag, J. A. et al. Age-Related Central Gain with degraded neural synchrony in the auditory brainstem of mice and humans. *Neurobiol. Aging.* **115**, 50–59 (2022).
17. Harris, K. C. et al. Afferent loss, GABA, and Central Gain in older adults: associations with speech recognition in noise. *J. Neurosci.* **42**, 7201–7212 (2022).
18. Kuwada, S. et al. Sources of the scalp-recorded amplitude-modulation following response. *J. Am. Acad. Audiol.* **13**, 188–204 (2002).
19. Shaheen, L. A., Valero, M. D. & Liberman, M. C. Towards a diagnosis of Cochlear Neuropathy with Envelope following responses. *J. Assoc. Res. Otolaryngol.* https://doi.org/10.1007/s10162-015-0539-3 (2015).
20. Joris, P. X., Schreiner, C. E. & Rees, A. Neural processing of amplitude-modulated sounds. *Physiol. Rev.* **84**, 541–577 (2004).
21. Herdman, A. T. et al. Intracerebral sources of human auditory steady-state responses. *Brain Topogr.* **15**, 69–86 (2002).
22. Picton, T. W., John, M. S., Dimitrijevic, A. & Purcell, D. Human auditory steady-state responses. *Int. J. Audiol.* **42**, 177–219 (2003).
23. Parthasarathy, A. & Bartlett, E. Two-channel recording of auditory-evoked potentials to detect age-related deficits in temporal processing. *Hear. Res.* **289**, (2012).
24. Pichora-Fuller, M. K. et al. Hearing impairment and cognitive energy: the Framework for understanding Effortful listening (FUEL). *Ear Hear.* **37**, 5S–27S (2016).
25. Kahneman, D. & Beatty, J. Pupil diameter and load on memory. *Science* **154**, 1583 (1966).
26. Peelle, J. E. Listening effort: how the Cognitive consequences of Acoustic Challenge are reflected in brain and behavior. *Ear Hear.* **39**, 204–214 (2018).

27. Beatty, J., Phasic not tonic pupillary responses vary and with auditory vigilance performance. *Psychophysiology* **19**, 167–172 (1982).
28. Tun, P. A., McCoy, S., Wingfield, A. & Aging Hearing acuity, and the attentional costs of Effortful listening. *Psychol. Aging*. **24**, 761–766 (2009).
29. Piquado, T., Isaacowitz, D. & Wingfield, A. Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* **47**, 560–569 (2010).
30. McGinley, M. J., David, S. V. & McCormick, D. A. Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron* **87**, 179–192 (2015).
31. Winn, M. B., Edwards, J. R. & Litovsky, R. Y. The impact of Auditory Spectral Resolution on listening Effort revealed by Pupil Dilation. *Ear Hear.* **36**, e153–e165 (2015).
32. Reimer, J. et al. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* **7**, 13289 (2016).
33. McHaney, J. R., Tessmer, R., Roark, C. L. & Chandrasekaran, B. Working memory relates to individual differences in speech category learning: insights from computational modeling and pupillometry. *Brain Lang.* **222**, 105010 (2021).
34. Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J. & Banerjee, S. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* **116**, 2395–2405 (2004).
35. Galbraith, G. et al. Murine auditory brainstem evoked response: putative two-channel differentiation of peripheral and central neural pathways. *J. Neurosci. Methods*. **153**, 214–220 (2006).
36. Ping, J. L. et al. Auditory evoked responses in the rat: transverse mastoid needle electrodes register before cochlear nucleus and do not reflect later inferior colliculus activity. *J. Neurosci. Methods*. **161**, 11–16 (2007).
37. Wang, L., Bharadwaj, H. & Shinn-Cunningham, B. Assessing cochlear-place specific temporal coding using Multi-band Complex tones to measure envelope-following responses. *Neuroscience* **407**, 67–74 (2019).
38. Encina-Llamas, G., Dau, T. & Epp, B. On the use of envelope following responses to estimate peripheral level compression in the auditory system. *Sci. Rep.* **11**, 6962 (2021).
39. Mirman, D. Growth Curve Analysis and Visualization Using R. *Routledge & CRC Press* (2014). https://www.routledge.com/Growth-Curve-Analysis-and-Visualization-Using-R/Mirman/p/book/9781466584327
40. Bharadwaj, H. M., Masud, S., Mehraei, G., Verhulst, S. & Shinn-Cunningham, B. G. Individual Differences Reveal Correlates of Hidden Hearing Deficits. *J. Neurosci.* **35**, 2161–2172 (2015).
41. Watkins, K. E., Strafella, A. P. & Paus, T. Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* **41**, 989–994 (2003).
42. Shinn-Cunningham, B. G. & Best, V. Selective attention in normal and impaired hearing. *Trends Amplif.* **12**, 283–299 (2008).
43. Rönnberg, J., Rudner, M., Lunner, T. & Zekveld, A. A. When cognition kicks in: working memory and speech understanding in noise. *Noise Health*. **12**, 263–269 (2010).
44. Golestani, N., Hervais-Adelman, A., Obleser, J. & Scott, S. K. Semantic versus perceptual interactions in neural processing of speech-in-noise. *Neuroimage* **79**, 52–61 (2013).
45. Zekveld, A. A., Rudner, M., Johnsrude, I. S. & Rönnberg, J. The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* **134**, 2225–2234 (2013).
46. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc. Natl. Acad. Sci. U S A*. **111**, 7126–7131 (2014).
47. Du, Y., Buchsbaum, B. R., Grady, C. L. & Alain, C. Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat. Commun.* **7**, 12241 (2016).
48. McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E. & Munro, K. J. Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology* **54**, 193–203 (2017).
49. Kousaie, S. et al. Language learning experience and mastering the challenges of perceiving speech in noise. *Brain Lang.* **196**, 104645 (2019).
50. Price, C. N. & Bidelman, G. M. Attention reinforces human corticofugal system to aid speech perception in noise. *NeuroImage* **235**, 118014 (2021).
51. Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E. & Johnsrude, I. S. Attentional Modulation of Envelope-Following Responses at Lower (93–109 Hz) but Not Higher (217–233 Hz) Modulation Rates. *JARO* 19, 83–97 (2018).
52. Shinn-Cunningham, B. Cortical and Sensory Causes of Individual Differences in selective attention ability among listeners with normal hearing thresholds. *J. Speech Lang. Hear. Res.* **60**, 2976–2988 (2017).
53. DiNino, M., Holt, L. L. & Shinn-Cunningham, B. G. Cutting through the noise: noise-Induced Cochlear Synaptopathy and Individual Differences in Speech understanding among listeners with normal audiograms. *Ear Hear.* **43**, 9–22 (2022).
54. Mamo, S. K. & Helfer, K. S. Speech understanding in modulated noise and Speech maskers as a function of cognitive status in older adults. *Am. J. Audiol.* **30**, 642–654 (2021).
55. Xie, Z., Zinszer, B. D., Riggs, M., Beevers, C. G. & Chandrasekaran, B. Impact of depression on speech perception in noise. *PLoS One*. **14**, e0220928 (2019).
56. Smayda, K. E., Engen, K. J. V., Maddox, W. T. & Chandrasekaran, B. Audio-Visual and Meaningful Semantic Context Enhancements in older and younger adults. *PLOS ONE*. **11**, e0152773 (2016).
57. Grant, K. J. et al. Predicting neural deficits in sensorineural hearing loss from word recognition scores. *Sci. Rep.* **12**, 8929 (2022).
58. Holmes, E. & Griffiths, T. D. Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. *Sci. Rep.* **9**, 16771 (2019).
59. Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J. C. & Maiste, A. C. Potentials-evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J. Acoust. Soc. Am.* **82**, 165–178 (1987).
60. Boettcher, F. A., Poth, E. A., Mills, J. H. & Dubno, J. R. The amplitude-modulation following response in young and aged human subjects. *Hear. Res.* **153**, 32–42 (2001).
61. He, N. J., Mills, J. H., Ahlstrom, J. B. & Dubno, J. R. Age-related differences in the temporal modulation transfer function with pure-tone carriers. *J. Acoust. Soc. Am.* **124**, 3841–3849 (2008).
62. Ruggles, D., Bharadwaj, H. & Shinn-Cunningham, B. G. Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15516–15521 (2011).
63. Dimitrijevic, A. et al. Human envelope following responses to Amplitude Modulation: effects of aging and modulation depth. *Ear Hear.* **37**, E322–E335 (2016).
64. Chandrasekaran, B. & Kraus, N. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* **47**, 236–246 (2010).
65. Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S. & Zatorre, R. J. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* **7**, 11070 (2016).
66. Coffey, E. B. J. et al. Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* **10**, 5036 (2019).
67. Bidelman, G. M. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* **175**, 56–69 (2018).
68. Gnanateja, G. N. et al. Frequency-Following Responses to Speech Sounds Are Highly Conserved across Species and Contain Cortical Contributions. *eNeuro* 8, ENEURO.0451-21.2021 (2021).

69. Kiren, T. et al. The generator of amplitude-modulation following response. *Acta Otolaryngol.* 28–33 (1994).
70. Mepani, A. M. et al. Envelope following responses predict speech-in-noise performance in normal-hearing listeners. *J. Neurophysiol.* **125**, 1213–1222 (2021).
71. Galbraith, G. C., 2-channel brain-stem frequency-following responses to pure-tone & and missing fundamental stimuli. *Electroencephalogr. Clin. Neurophysiol.* **92**, 321–330 (1994).
72. Galbraith, G. C. et al. Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci. Lett.* **292**, 123–127 (2000).
73. King, A., Hopkins, K. & Plack, C. J. Differential Group Delay of the frequency following response measured vertically and horizontally. *Jaro-Journal Association Res. Otolaryngol.* **17**, 133–143 (2016).
74. Smith, M. L., Winn, M. B. & Fitzgerald, M. B. A large-scale study of the relationship between degree and type of hearing loss and Recognition of Speech in quiet and noise. *Ear Hear.* **45**, 915–928 (2024).
75. Cowan, T. et al. Masked-Speech Recognition for linguistically diverse populations: a focused review and suggestions for the future. *J. Speech Lang. Hear. Res.* **65**, 3195–3216 (2022).
76. Wild, C. J. et al. Effortful listening: the Processing of degraded Speech depends critically on attention. *J. Neurosci.* **32**, 14010–14021 (2012).
77. Kuchinsky, S. E. et al. Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology* **50**, 23–34 (2013).
78. Kuchinsky, S. E. et al. Task-related vigilance during Word Recognition in noise for older adults with hearing loss. *Exp. Aging Res.* **42**, 50–66 (2016).
79. Winn, M. B. Rapid Release from listening effort resulting from semantic context, and effects of Spectral Degradation and Cochlear implants. *Trends Hear.* **20**, (2016).
80. McLaughlin, D. et al. (ed, J.) Give me a break! Unavoidable fatigue effects in cognitive pupillometry. *Psychophysiology* **e14256** https://doi.org/10.1111/psyp.14256 (2023).
81. Aston-Jones, G. & Cohen, J. D. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* **28**, 403–450 (2005).
82. de Gee, J. W. et al. Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. *Elife* **9**, e54014 (2020).
83. McGinley, M. J. et al. Waking state: Rapid variations modulate neural and behavioral responses. *Neuron* **87**, 1143–1161 (2015).
84. Ohlenforst, B. et al. Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear. Res.* **351**, 68–79 (2017).
85. Zink, M. E. et al. Increased listening effort and cochlear neural degeneration underlie behavioral deficits in speech perception in noise in normal hearing middle-aged adults. 08.01.606213 Preprint at (2024). https://doi.org/10.1101/2024.08.01.606213 (2024).
86. Hunter, L. L. et al. Extended high frequency hearing and speech perception implications in adults and children. *Hear. Res.* **397**, 107922 (2020).
87. Zadeh, L. M. et al. Extended high-frequency hearing enhances speech perception in noise. *PNAS* https://doi.org/10.1073/pnas.1903315116 (2019).
88. Monson, B. B., Rock, J., Schulz, A., Hoffman, E. & Buss, E. Ecological cocktail party listening reveals the utility of extended high-frequency hearing. *Hear. Res.* **381**, 107773 (2019).
89. Akeroyd, M. A. 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. *Int. J. Audiol.* **47** (Suppl 2), S53–71 (2008).
90. CHABA. Speech understanding and aging. *J. Acoust. Soc. Am.* **83**, 859–895 (1988).
91. Yi, H., Smiljanic, R. & Chandrasekaran, B. The Effect of Talker and Listener depressive symptoms on Speech Intelligibility. *J. Speech Lang. Hear. Res.* **62**, 4269–4281 (2019).
92. Chandrasekaran, B., Van Engen, K., Xie, Z., Beevers, C. G. & Maddox, W. T. Influence of depressive symptoms on speech perception in adverse listening conditions. *Cogn. Emot.* **29**, 900–909 (2015).
93. Tyler, R. S. & Baker, L. J. Difficulties experienced by tinnitus sufferers. *J. Speech Hear. Disord.* **48**, 150–154 (1983).
94. Vielsmeier, V. et al. Speech Comprehension difficulties in Chronic Tinnitus and its relation to Hyperacusis. *Front. Aging Neurosci.* **8**, 293 (2016).
95. Ivansic, D. et al. Impairments of Speech Comprehension in patients with Tinnitus—A Review. *Front. Aging Neurosci.* **9**, (2017).
96. Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
97. Beck, A. T., Steer, R. A. & Carbin, M. G. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. *Clin. Psychol. Rev.* **8**, 77–100 (1988).
98. Wilson, P. H., Henry, J., Bowen, M. & Haralambous, G. Tinnitus reaction questionnaire: psychometric properties of a measure of distress associated with tinnitus. *J. Speech Hear. Res.* **34**, 197–201 (1991).
99. Winn, M. B., Wendt, D., Koelewijn, T. & Kuchinsky, S. E. Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear.* **22**, 2331216518800869 (2018).
100. McHaney, J. R., Schuerman, W. L., Leonard, M. K. & Chandrasekaran, B. Low amplitude transcutaneous auricular vagus nerve stimulation modulates performance but not pupil size during non-native speech category learning. 07.19.500625 Preprint at (2022). https://doi.org/10.1101/2022.07.19.500625 (2022).
101. McHaney, J. R., Schuerman, W. L., Leonard, M. K. & Chandrasekaran, B. Transcutaneous Auricular Vagus nerve stimulation modulates performance but not pupil size during nonnative Speech Category Learning. *J. Speech Lang. Hear. Res.* **66**, 3825–3843 (2023).
102. R Core Team. R: a language and environment for statistical computing. (2022). https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing
103. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
104. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: tests in Linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
105. Benjamini, Y. & Hochberg, Y. Controlling the false Discovery rate: a practical and powerful Approach to multiple testing. *J. Roy. Stat. Soc.: Ser. B (Methodol.)*. **57**, 289–300 (1995).
106. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* (Springer, 2002). https://doi.org/10.1007/978-0-387-21706-2
107. Erdfelder, E., Faul, F. & Buchner, A. GPOWER: a general power analysis program. *Behav. Res. Methods Instruments Computers.* **28**, 1–11 (1996).

Acknowledgements

This work was supported by NIH P50DC015817 (DBP), NIH R21DC018882 (AP), T32DC011499 (trainee: JRM), and F31DC020085 (JRM). The authors thank Jessica L'Heureux and Jennifer Klara for administrative and technical support.

Author contributions

Conceptualization, AP and DBP; Methodology, AP, DBP, KEH, JRM; Software, KEH; Formal Analysis, JRM; Investigation, AP; Resources, DBP and AP; Data Curation, JRM; Writing – Original Draft, JRM and AP; Writing – Review & Editing, JRM, AP, KEH, and DBP; Visualization, JRM and AP; Supervision, AP; Project Administration, AP; Funding Acquisition, DBP and JRM.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-81673-8.

Correspondence and requests for materials should be addressed to A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.