



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/240108/>

Version: Accepted Version

---

**Article:**

Whitley, A., Corps, R.E. and Hadley, L.V. (2026) Reduced intelligibility explains delayed responses to questions in older adults with hearing loss. *Language, Cognition and Neuroscience*, 41 (4). pp. 548-559. ISSN: 2327-3798

<https://doi.org/10.1080/23273798.2026.2640963>

---

© 2026 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in *Language, Cognition and Neuroscience* is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# **Reduced intelligibility explains delayed responses to questions in older adults with hearing loss**

Alexina Whitley<sup>a\*</sup>, Ruth E. Corps<sup>b</sup>, and Lauren V. Hadley<sup>a</sup>

<sup>a</sup>*Hearing Sciences – Scottish Section, University of Nottingham, Glasgow, UK;*

<sup>b</sup>*School of Psychology, University of Sheffield, Sheffield, UK*

e-mail: alexina.whitley@nottingham.ac.uk

School of Medicine,

University of Nottingham,

Hearing Sciences – Scottish Section,

New Lister Building Level 3, Glasgow Royal Infirmary,

10-16 Alexandra Parade, Glasgow G31 2ER

## **Funding Details**

This work was supported by a UKRI Future Leaders Fellowship [grant number UKRI1059], and the Economic and Social Research Council [grant number ES/X001148/1].

## **Disclosure of interest**

The authors report no conflict of interest

## **Data availability statement**

Data available on request from the authors

## **Reduced intelligibility explains delayed responses to questions in older adults with hearing loss**

Typical conversation is characterized by tightly coordinated turn-taking, meaning that listeners must begin planning their response before their partner has finished talking. But concurrently planning and comprehending incoming speech is resource intensive. This study investigates the effects of intelligibility (by manipulating stimulus presentation level) and hearing loss (through comparison of older adults with normal vs impaired hearing) on the timing of response planning using a verbal question answering task. When information critical for response planning occurred early in the question, participants were faster to answer than when it occurred late, indicating an early planning approach. Although hearing impaired participants showed slower responses than normal hearing participants when listening to conversation-level speech, they showed no change in planning strategy. Furthermore, when intelligibility was equated, groups performed similarly, suggesting that the turn-taking delays that have been reported for hearing impaired interlocutors may be driven by reduced intelligibility impacting planning.

Keywords: hearing loss; response planning; conversation timing; language production

## **Introduction**

Conversation is typically characterized by tightly coordinated turn-taking, whereby speakers rapidly exchange utterances with minimal gaps or overlaps. In English, the average gap between one speaker finishing their turn and the next speaker beginning is around 200 milliseconds (Stivers et al., 2009), yet the time required to plan a single word is considerably longer (approximately 600 milliseconds, Indefrey & Levelt, 2004, and 1500ms for a simple sentence, Griffin & Bock, 2000). Thus, speakers cannot wait until the end of a conversational turn before beginning to plan their response, instead, they must begin planning their response before their partner has finished (Bögels et al., 2015; Corps et al., 2018). Although theories of turn-taking agree that interlocutors manage conversation demands by planning responses early, i.e., while still comprehending the current speaker's turn, evidence supporting such accounts come from populations that find speech processing relatively undemanding (such as younger adults). In this paper, we build on this prior work to investigate the timecourse of response planning in people listening under demanding conditions. Specifically, we investigate older adult listeners with normal hearing vs hearing loss, the latter of whom experience increased cognitive demand due to effortful listening and may not have cognitive resources available to support such early planning.

## ***Response planning***

A key question is how soon do listeners start planning their response once they have enough information to do so. The 'early planning hypothesis' proposes that listeners plan their responses as early as possible, and then hold their response in mind while they wait for turn-end cues that signal the appropriate time to speak (Levinson & Torreira, 2015). According to this account, listeners simultaneously plan and buffer a

linguistic response while still comprehending incoming speech input. Numerous studies have found evidence consistent with such early planning. Bögels et al. (2015) used an interactive question answering paradigm, measuring response times and EEG during a question answering task, in which critical information necessary to prepare an answer was available either early in the question (i.e., “What animal barks and has four legs?”), or late (i.e., “What animal has four legs and barks?”). Their findings showed that participants were both faster to respond and had greater activation in brain areas related to speech planning when critical content appeared early, suggesting that people initiate response planning as soon as relevant information is available. Furthermore, Corps et al. (2018) asked participants to verbally respond to yes/no questions in which the final word was either highly predictable, enabling early response planning (i.e., “Are dogs your favourite animal?”), or unpredictable, necessitating later planning (i.e., “Have you ever been to the city of Paris?”). Participants again responded more quickly to questions where the final word was predictable, indicating that response planning began as soon as possible regardless of distance from the turn-end. Other consistent results have been found using a variety of alternative paradigms (e.g., Barthel et al., 2016 measured pupillometry and response latencies during a list-completion task).

But despite the above evidence for early planning, numerous studies have demonstrated that speech planning and speech comprehension are related cognitive processes that rely on overlapping neural circuits (Kemper et al., 2003; Kubose et al., 2006; Segaert et al., 2012) . Holding a response in mind may therefore interfere with concurrent comprehension of the incoming turn (Bögels et al., 2018) and require additional cognitive resources (Barthel & Sauppe, 2019; Fairs et al., 2018). Indeed, studies have shown that planning responses while concurrently comprehending incoming speech is associated with higher processing load than planning in silence, as

measured by the pupillary response (Barthel & Sauppe, 2019). Thus, postponing response planning could reduce this interference.

The ‘late planning hypothesis’ is an alternative account that posits that listeners only begin the response planning process when they know that the speaker will soon reach the end of their turn (i.e., by detecting turn-end cues), even if they may have had the information and understanding to begin response planning earlier (Sjerps & Meyer, 2015). Consistent with this account, Sjerps and Meyer (2015) asked participants to complete a finger tapping task whilst they alternated between listening to pre-recorded utterances and describing rows of pictures. Participants’ finger tapping performance decreased only shortly before the end of the incoming speech (and thus only shortly before beginning their own speech turn). Eye-tracking data further revealed that looks to the pictures requiring description began only shortly before the finger tapping decline. Together these findings suggest that response planning was delayed until close to the speaking time, differing from many of the studies discussed above. A potential explanation for such contradictory findings lies in the demands of the task: the Sjerps and Meyer paradigm required cognitive resources to be distributed across the dual tapping and turn-taking tasks, unlike the pure tasks described previously. Hence this work suggests that in highly demanding situations, listeners may indeed delay speech planning and show a later planning pattern of behaviour.

### ***Hearing loss and response planning***

People with hearing loss experience increased listening demand, and show both slower and more variable turn-timing in conversation (Petersen et al., 2022; Sørensen et al., 2020). It is therefore possible that hearing loss leads to delays in speech planning. One explanation for turn-taking delays in people with hearing loss is that they find listening

particularly effortful, which increases the cognitive demands of speech processing (e.g., Peelle, 2018; Pichora-Fuller et al., 2016); consequently reducing resources for other higher-level processing (see Fernandez et al., 2024 for discussion). As planning a response while simultaneously listening is demanding (e.g., Barthel & Sauppe, 2019) and requires additional cognitive resources, early planning may therefore not be feasible for people with hearing loss, and planning may be delayed. Importantly, such delays need not reflect a voluntary strategy; they may arise either through deliberate postponement to manage cognitive load or through an implicit/learned adaptation to listening difficulty. Additionally, if people with hearing loss take longer to process incoming speech more broadly (Wendt et al., 2015), less time would be available for planning responses in the first place. Hence early response planning may be too cognitively demanding, or slow, for people with hearing loss, and thus a late response planning approach could underlie the longer and more variable response latencies reported for conversational turn taking.

### ***Current Study***

In the current study we examine how cognitive demand affects the timing of response planning in two groups of older adult listeners: one with normal hearing, and one with impaired hearing. Several previous studies have demonstrated that people use the content of incoming speech to begin planning responses early. However, these studies have primarily focused on young adult listeners in ideal listening conditions. Hearing loss has been shown to affect response timing in conversation, which could be due to increased listening demand leading to delays in response planning. This study therefore uses the question answering paradigm of Bögels et al. (2018) to investigate how response planning is affected by challenging listening conditions. Specifically, we

address the effect of hearing loss between normal hearing and hearing-impaired groups, as well as the effect of speech intelligibility within participants. We hypothesised that when cognitive resources are allocated to effortful listening, response planning would be delayed. Importantly, this delay need not reflect a deliberate decision to “wait”; it may arise because increased demand postpones the point at which the question is sufficiently understood and an answer is available for formulation. Importantly, hearing loss may affect response planning due to broadly altering the cognitive processes involved in listening, or due to reducing speech intelligibility in a way that could be replicated in normal hearing listeners. Thus, we compare two older adult groups listening at a conversational speech level of 60 dB, as well as at individually determined levels selected to ensure matched intelligibility scores (outlined below). This design thus allows us to address both the direct effect of hearing loss at natural conversational levels, and the more general effect of reduced intelligibility on response planning.

## Methods

### *Participants*

The study included 25 normal hearing (NH, 15 female, 10 male,  $age_{mean} = 63.92$ ,  $age_{sd} = 6.68$ ) and 28 hearing impaired (HI, 19 female, 9 male,  $age_{mean} = 69.75$ ,  $age_{sd} = 7.02$ ) older adults. The NH participants had better-ear pure-tone average hearing thresholds at or below 25 dB HL across 0.5, 1, 2, and 4kHz (mean = 16.98, SD = 5.14). The HI participants had better-ear pure-tone average hearing thresholds at or above 35 dB HL across 0.5, 1, 2 and 4kHz (mean = 52.85, SD = 7.82). Audiograms for both groups are plotted in Figure 1. All participants were recruited from the Glasgow region and were fluent in English. Participant demographics are shown in Table 1.

Ethical approval was obtained from the University of Nottingham Faculty of Medicine and Health Science Research Ethics Committee (REC reference: FMHS 423-1221). Participants provided written informed consent to participate in this study.

[Figure 1 here]

[Table 1 here]

### *Stimuli*

Stimuli comprised 120 general knowledge question pairs. Questions were constructed to provide both ‘general’ information to narrow potential answers to a category (e.g., common household pet), and critical information to constrain potential answers to a single response (e.g., barks). Each item pair included one version in which the critical content was presented EARLY in the sentence (e.g., “Which animal barks and is also a common household pet?”), and one in which the critical content was presented LATE in the sentence (e.g., “Which animal is a common household pet and also barks?”). In the

EARLY version, answers could be planned early or late depending on strategy, whereas in the LATE version, answers could only be planned late.

Stimuli were selected based on an online pre-test of 20 participants (10 female,  $age_{mean} = 65.35$   $age_{sd} = 4.93$ ), recruited via Prolific ([www.prolific.com](http://www.prolific.com)). All participants were normal hearing native English speakers and were of Scottish nationality. An initial set of 155 question pairs were divided into two stimulus lists, each containing one version of each question and an equal number of EARLY- and LATE- critical content versions. Participants were randomly assigned to one of the two lists (10 participants per list). They were required to read each question and “Answer each question to the best of your ability. This can be a single word, or it can be longer and contain any number of words”. Participants typed their responses.

To reduce stimulus differences based on difficulty of selecting the question’s answer, the similarity of question answers was assessed using Latent Semantic Analysis (LSA) matrix comparisons with the general reading corpus provided by Colorado University Boulder (<http://lsa.colorado.edu>). LSA provides a measure of the degree of similarity between responses by analysing how often these words appear in the same contexts. LSA values range from 1, indicating that the different responses occur in identical contexts, to 0, indicating that the different responses occur in different contexts. Question LSA was calculated by averaging the LSA values for all pairwise comparisons between answers. Of 155 stimuli pairs, 35 were removed (4 due to being very similar to other questions in the list, as well as the 31 with the lowest LSA values), to create a final experimental list of 120 questions. Average question LSA values for the 120 question pairs did not significantly differ between EARLY ( $questionLSA_{mean} = 0.94$ ) and LATE ( $questionLSA_{mean} = 0.94$ ) stimuli ( $p = 0.36$ ). Thus, participants tended

to agree on the answers regardless of whether stimuli were presented in the EARLY or LATE version.

Question stimuli were recorded with a male Scottish accented speaker using a Blue YETI USB microphone at a sampling rate of 48kHz. Question durations were between 2731ms and 7580s. The duration of questions in the EARLY condition (mean = 46218ms) did not significantly differ from their pair in the LATE condition (mean = 4582ms;  $p = .098$ ).

### ***Other measures***

Processing speed was measured using the digit symbol substitution test (DSST; Wechsler, 1981). Participants were presented with a key, pairing unique symbols with digits (1–9). Below the key was a test section containing rows of numbers with empty boxes. Participants were required to write the corresponding symbols into the spaces below the numbers as quickly and accurately as possible for 90 seconds.

### ***Design and Procedure***

The experiment had a 2 (critical content EARLY/LATE) by 2 (presentation level 60 dB/Adaptive) by 2 (participant group HEARING LOSS/NORMAL HEARING) design. The experiment was divided into two blocks. In one block the questions were presented at a fixed level of 60 dB. In the other block the questions were presented at an individually-determined adaptive threshold. Order of blocks was counterbalanced between participants. Critical content was balanced within each presentation level (i.e., there were 30 trials per combination of critical content and presentation level conditions) and was counterbalanced such that participants heard only one version of each question.

Prior to beginning the study, participants with hearing aids were instructed to remove them. Participants first underwent an audiogram and the adaptive track procedure to determine the level for the adaptive presentation level condition (to equate intelligibility across individuals and groups). Based on participants' designation as NH or HI from their audiogram, different starting levels for the adaptive track were selected. Speech was always presented in quiet. For normal hearing listeners the start level was set at 50 dB and for hearing impaired listeners it was set at 70 dB. Participants were then presented with a series of spoken questions, similar to those in the main experiment. Sentences ranged from 5-8 words (mean = 6.6), which was slightly shorter than the stimuli used in the main experiment to reduce demands on working memory, and included sentences such as: "Which temperature does water boil at?", and "What colour is a New York taxi?". After each stimulus participants were asked to repeat back all the words in the question. Responses were scored as correct only if all words in the sentence were repeated accurately and in the correct order. Any omitted, substituted, or incorrectly repeated word resulted in the sentence being scored as incorrect. Sentence presentation level was then adjusted adaptively using a 2-down, 1-up staircase procedure. If the participant responded correctly for two consecutive trials, the presentation level was decreased, and if they responded incorrectly for one trial, the presentation level was increased. A reversal was defined as a point at which the direction of the level adjustment changed, indicating a shift from an increase to a decrease in presentation level, or vice versa. The level was changed in steps of 6 dB for the first 4 reversals, and steps of 3 dB for 5 reversals. A maximum of 40 trials were presented. We used a 71% adaptive threshold (corresponding to the level at which listeners could correctly identify approximately 71% of the stimuli) estimated by averaging the presentation level across the last 5 reversals (Levitt, 1971). This section

ended with a further five short sentences at the 60 dB level, and at the selected adaptive presentation level, to get final intelligibility scores (see Table 1).

Participants then began the main experiment. As in (Corps et al., 2018; 2024), participants were instructed at the start of each block to “Answer the question with the word or words that you think are most appropriate as quickly as possible. Do not wait until the speaker has finished the question and has stopped speaking. Instead, you should answer as soon as you expect the speaker to finish the question.” At the start of each trial a fixation cross appeared for 500ms, then turned red from when the question began until 2000ms after the question ended (considered the timeout point for responses). Then a blank screen appeared for 500ms before the next trial began. Participants’ verbal responses were recorded via a dynamic cardioid microphone. Participants first completed five main experiment practice trials at 60 dB to familiarise themselves with the task before beginning the first block. The experiment was presented in MATLAB (2021a), blocked by presentation level, with order counterbalanced between participants. Questions in each block were presented in an individually randomised order. After the main experiment participants completed the DSST (Wechsler, 1981).

### ***Data analysis***

Verbal responses were recorded from the onset of the question to 2000ms following the question end. The onset of participant responses was calculated automatically in MATLAB (2021a) using an automatic voice detection threshold. Each trial was manually checked to ensure the onset of the response was correct (i.e., to ensure non-speech sounds such as breaths or movement hadn’t been detected as the onset of the response). Verbal response times were defined as the difference between the question

end and the beginning of the participant response, ignoring any manually identified non-speech sounds. Thus, negative response times indicate that participants responded before the end of the question, and positive response times indicate that participants responded after the question end. Prior to statistical analysis, we removed trials where (1) participants did not produce a response within 2000ms (26.57%), (2) we could not determine what the participant said (and thus could not ensure it was a relevant answer; 0.35%), or (3) participants responded that they either did not know or did not hear the question (0.27%). Total proportion of removed trials was 27.19% (see Table 2 for summary of percentage of removed data by hearing group and presentation level). We also coded the accuracy of responses. Answers were considered valid if they matched the answer produced in the pre-test, but we also accepted other question-relevant answers (even if they were not correct - e.g., 1945 instead of 1939 in response to “Which year was in the 1900s and was the start of World War Two?”). Incorrect responses were a small subset of the retained data (137 trials; 2.15%). In sum, all interpretable, question-relevant responses were retained for analyses regardless of accuracy.

[Table 2 here]

Response times were analysed with linear mixed-effects models using the `lmer` function of the `lme4` package in RStudio. Presentation level in the adaptive block was always lower than 60 dB for NH participants, but varied for HI participants (since for some HI participants, their intelligibility at 60 dB was still very poor). As such, three separate models were fitted. To assess how intelligibility affected response planning, one model was fit for NH participants comparing 60 dB vs the adaptive presentation levels. To assess how HL affects response planning at conversational levels, one model was fit across participant groups at the 60 dB presentation level. Finally, to assess how

HL affects response planning when intelligibility was equated between groups, one model was fit across participant groups at the adaptive presentation level. Main effects of critical content (reference level: EARLY vs. LATE), presentation level (reference level: 60 dB vs. Adaptive), and hearing group (reference level: NH vs. HI) were contrast coded (-0.5, 0.5). As previous research suggests that longer questions tend to elicit earlier responses (e.g., Corps & Pickering, 2024) we also included question duration as a main effect. Additionally, we included a measure of processing speed as prior work has shown that faster processing speed supports anticipatory language processing, which may similarly facilitate earlier response planning (Huettig & Janse, 2016). By participant random effects were included in all models, and by item random effects were included in response time accuracy models.

To preview our results, we found the following null effects: (1) no difference between 60 dB and adaptive presentation levels in NH, and no interaction with critical content; (2) no interaction between hearing group and critical content when at the 60 dB level; and (3) no difference between NH and HI participants at the adaptive level, and no interaction with critical content. We calculated Bayes Factors for all predictors by fitting Bayesian mixed effects models using the brms package (version 2.22.0). We calculated Bayes Factors by comparing a model with the predictor of interest (e.g., critical content; M1) to a reduced model without this predictor (M0). A Bayes Factor of approximately 1 indicates no evidence in favour of either model. As the Bayes Factor increases over 3, evidence in favour of the M1 strengthens; as the Bayes Factor decreases under 0.3, evidence strengthens in favour of the M0 (e.g., Kass & Raftery, 1995; Lee & Wagenmakers, 2014).

We fitted models using informative priors, based on our expectation from previous studies. All priors were fitted using a normal distribution. Based on Corps and

Pickering (2024), who used a similar EARLY/LATE manipulation with young adults without hearing loss, we expected answer times to average around 400ms, with some variability, and so for the intercept we set a prior with a mean of 400ms and a standard deviation of 200ms. We expected a negative effect of critical content (i.e., faster responses for EARLY than LATE question types) and we expected this effect to be around 600ms based on Corps and Pickering (2024) Thus, we set a prior with a mean of -300ms and a standard deviation of 200ms for the main effect of critical content.

For the NH analysis, we did not have any previous data to estimate the difference between the 60 dB and adaptive presentation levels. However, we expected participants to be faster at the 60 dB than the adaptive presentation level, given that the adaptive level was the lowest presentation level at which they could achieve 71% intelligibility and should thus have been more cognitively demanding. As a result, we set a prior with a mean of -100ms and a standard deviation of 200ms for the main effect of presentation level. We did not expect an interaction between critical content and presentation level, and so we set a prior with a mean of 0ms and a standard deviation of 200ms for the interaction coefficient.

For the second and third models, comparing NH and HI participants at the 60 dB and adaptive presentation levels, we estimated a group difference of around 200ms based on Corps et al. (2025) who investigated response planning in people with NH, people with HL listening at high-demand (adaptive), and people with HL listening at low-demand (fixed). As a result, we set a prior with a mean of -100ms and a standard deviation of 200 ms for the main effect of hearing group. In both models, we did not expect an interaction between hearing group and critical content, and so we set a prior with a mean of 0ms and a standard deviation of 200ms for the interaction coefficient.

In all models, we fit the standard deviation parameter to have a mean of 0ms and a standard deviation of 50ms and sigma to have a mean of 0ms and a standard deviation of 100ms (Corps & Pickering, 2024). We did not calculate Bayes Factors for question duration or DSST score because we included these predictors as control variables; we were not interested in whether they affected response times on their own.

We first fitted a model that simulated the data from the actual priors, and then visualized the distribution of effects to ensure they matched our expectations. Once we confirmed that the priors seemed plausible, we fitted models with the actual data. Bayes Factors are sensitive to the choice of prior, and so we also conducted a sensitivity analysis (Schad et al., 2023). We kept the same means as defined in our informative priors, but we changed the standard deviation of each parameter. In particular, we defined a range of priors with standard deviations from 300 to 1000ms in increments of 100ms, representing increasingly looser priors and increasing uncertainty about the effect.

For each predictor, we report coefficient estimates (b), standard errors (SE), and t-values for each predictor. We assume that a t-value of 1.96 or greater indicates significance at the 0.05 alpha level (Baayen et al., 2008). For Bayesian analysis, we report Bayes Factors (BF) from the informative model only, but we report whether the BF was consistent across the sensitivity analysis.

## **Results**

[Figure 2 here]

### ***Effects of Intelligibility in Normal Hearing Listeners only (60 dB vs adaptive presentation level)***

On average, NH participants responded 229.53ms after the question end. They were

faster to respond when the critical content was EARLY ( $M = -221.38$ ) than when it was LATE ( $M = 564.06$ ,  $\beta = -794.27$ ,  $SE = 66.16$ ,  $t = -12.01$ ,  $BF > 100$ ), and were faster to respond when the presentation level was at 60 dB ( $M = 132.77$ ) than when it was at the adaptive presentation level ( $M = 344.82$ ,  $\beta = -248.42$ ,  $SE = 29.28$ ,  $t = -8.49$ ,  $BF > 100$ ). The BF consistently supported the alternative hypotheses over the null in the sensitivity analysis (all BFs  $> 100$ ). Average response times for each condition are shown in Figure 2. There was no interaction between critical content and presentation level ( $\beta = 11.167$ ,  $SE = 58.32$ ,  $t = 0.19$ ,  $BF = 0.08$ ), and the BF consistently supported the null hypothesis over the alternative in the sensitivity analysis (all BFs  $< 0.04$ ). As found in prior studies, there was also a significant effect of question duration ( $\beta = -190.65$ ,  $SE = 32.90$ ,  $t = 5.79$ ; Corps et al., 2018), and participants with higher DSST scores were faster to respond ( $\beta = -14.61$ ,  $SE = 5.67$ ,  $t = -2.57$ ).

### ***Differences between Normal Hearing and Hearing Impaired Listeners***

#### *60 dB presentation level*

Participants across groups were faster to respond to questions when the critical content appeared EARLY ( $M = -110.40$ ) than when it appeared LATE ( $M = 643.84$ ,  $\beta = -755.12$ ,  $SE = 65.27$ ,  $t = -11.57$ ,  $BF > 100$ ), and the BF showed consistent evidence for the alternative hypothesis over the null in the sensitivity analysis (all BFs  $> 100$ ). Furthermore, normal hearing participants were faster to respond overall than hearing impaired participants ( $NH = 132.77$ ,  $HI = 375.53$ ,  $\beta = -220.93$ ,  $SE = 90.88$ ,  $t = -2.43$ ,  $BF = 56.82$ ), and the BF consistently supported the alternative hypothesis over the null in the sensitivity analysis (all BFs  $> 11.01$ ). There was no interaction between hearing group and critical content ( $\beta = -46.52$ ,  $SE = 56.42$ ,  $t = -0.82$ ,  $BF = 0.10$ ). The BF showed consistent evidence for the null hypothesis over the alternative in the sensitivity

analysis (all BFs < 0.07) Again, there was a significant effect of question duration ( $\beta = -170.92$ ,  $SE = 32.42$ ,  $t = -5.27$ ), and participants with higher DSST scores were faster to respond ( $\beta = -11.68$ ,  $SE = 4.65$ ,  $t = -2.51$ ).

#### *Adaptive presentation level*

Participants across groups were faster to respond to questions when the critical content was EARLY ( $M = 28.64$ ) than when it was LATE ( $M = 791.06$ ,  $\beta = -740.59$ ,  $SE = 60.53$ ,  $t = -12.24$ ,  $BF > 100$ ). The BF showed consistent evidence for the alternative hypothesis over the null in the sensitivity analysis (all BFs > 100). However, in this condition of equated intelligibility there was no longer a difference between response times for normal hearing ( $M = 344.82$ ) and hearing-impaired participants ( $M = 434.05$ ;  $\beta = -32.48$ ,  $SE = 94.05$ ,  $t = -0.35$ ,  $BF = 0.39$ ). The BF showed consistent evidence for the null hypothesis over the alternative in the sensitivity analysis (all BFs < 0.27). There was also no interaction between hearing group and critical content ( $\beta = -108.23$ ,  $SE = 60.68$ ,  $t = -1.78$ ,  $BF = 0.10$ ), and the BF showed consistent evidence for the null hypothesis over the alternative in the sensitivity analysis (all BFs < 0.20). Again, there was a significant effect of question duration ( $\beta = -194.65$ ,  $SE = 30.26$ ,  $t = -6.43$ ), and participants with higher DSST scores were faster to respond ( $\beta = -15.74$ ,  $SE = 4.83$ ,  $t = -3.26$ ).

#### *Exploratory analysis of excluded trials*

We modelled the probability of trial exclusion (no response within 2000ms and/or “didn’t know/didn’t hear”) using mixed-effects logistic regression with random intercepts for participant and item. Results reflected those reported above for response times: For NH listeners, exclusion probability was higher at the adaptive presentation level than 60 dB presentation level ( $p < .001$ ). At the 60 dB presentation level,

exclusion probability was higher for HI than NH participants ( $p < .001$ ) and higher for LATE than EARLY items ( $p < .05$ ), with no interaction. At the adaptive presentation level, exclusion probability was higher for LATE than EARLY items ( $p = .05$ ), with no main effect of group and no interaction.

### ***Supplementary visualisation of response times relative to critical content***

To address whether response times were aligned to the point at which the answer became identifiable, we additionally recalculated response times relative to the offset of the critical content word (e.g., bark) rather than the end of the question. We report individual response-time distributions for EARLY and LATE trials in the Supplementary Materials, which highlight individual differences in response timing strategy.

## **Discussion**

### ***Overview***

The current study investigated whether hearing loss and speech intelligibility affect the ability to start planning responses early in a conversational question-answering task. Specifically, we examined how the timing of critical content (EARLY/LATE) and different listening conditions (60 dB/Adaptive) differentially affected the timing of responses for listeners with normal and impaired hearing. Consistent with previous research (Bögels et al., 2015; Corps, Gambi, et al., 2018), we found that all participants responded more quickly when critical content appeared EARLY in the question, regardless of intelligibility, suggesting that they initiated response planning earlier, even when listening was demanding. However, HI participants showed slower response times at a conversational listening level, indicating that hearing loss affects the efficiency of response planning, though the lack of interaction with critical content suggests that it does not affect the approach to response planning. Finally, manipulation of intelligibility within the NH participants was able to replicate the slowed response times of the HI participants, suggesting that such delays with hearing loss may be explained by reduced intelligibility rather than broader changes to strategy or higher-level cognition.

### ***Effect of hearing loss at conversational speech levels***

We found that verbal response latencies were significantly longer for older adults with hearing loss than those with normal hearing at conversational listening levels, aligning with reports of longer floor transfer times in real conversation (Petersen et al., 2022). While both NH and HI participants demonstrated anticipatory response planning (i.e., responding significantly faster when critical content was available EARLY), there

was an overall delay in response times observed in the hearing-impaired group. While this may reflect a slowing of the response planning process itself, it could also be explained by later initiation of response planning. For instance, there is evidence that hearing loss is associated with slower speech processing (e.g. Wendt et al., 2015), which could delay the point at which relevant information becomes available to initiate response planning. In this account, response planning may proceed at a similar rate once initiated, but begin later as a result of delayed speech processing.

Surprisingly, we saw no interaction between hearing loss and the availability of the critical information necessary for response planning (i.e., EARLY or LATE). Given the greater difficulty of the listening task for HI listeners, we had expected a potential change of strategy, from early to late planning. However, we found that regardless of hearing ability, listeners began planning their responses earlier for EARLY questions than the LATE questions.

### ***Effect of intelligibility on response planning***

Interestingly, at the adaptive listening level, we found no significant difference in response times between normal hearing and hearing-impaired participants. This suggests that when speech intelligibility is equated and thus presumably speech comprehension is similarly challenging, normal hearing listeners behave similarly to people with hearing loss. Indeed, normal hearing listeners were significantly slower to respond to general knowledge questions when listening at this reduced intelligibility level relative to when listening at conversational levels (being on average 209.76ms delayed), in spite of all analyses focusing only on those trials in which participants provided a valid response. These changes are consistent with delays seen in turn-taking timing for normal hearing listeners when adding background noise to conversations

(Petersen et al., 2022; median FTO difference = 69ms).

### ***Planning, hearing loss, and intelligibility***

Although overall response times were slower under more challenging listening conditions, due to either hearing loss or lower presentation levels, participants consistently responded more quickly when the critical content occurred EARLY in the question. Thus, it appears that even when comprehension is challenging listeners begin to prepare their response early, while the question is still unfolding.

Together our findings suggest that the slowing of verbal responses in people with hearing loss may be a simple result of reduced intelligibility. When listening is difficult, listeners need to allocate more cognitive resources to speech comprehension (e.g., Pichora-Fuller et al., 2016), leaving fewer resources available for concurrent response planning. In order to produce responses at the appropriate time, listeners must be able to comprehend and predict the end of the incoming turn, whilst at the same time planning a verbal response. However, when the speech signal is degraded or listening more demanding (for instance due to hearing loss, increased noise, or reduced presentation level), listeners may find it harder to match what was heard to the mental lexicon, and thus processing may be delayed. According to the ease of language understanding (ELU) model (Rönnberg et al., 2013), this could require additional cognitive resources to enable repair or allow inference of missing information from the speech signal. Therefore, cognitive demands may be increased, resulting in slower response times.

We might have expected a change of strategy, from early to late planning when listening demands were higher, either due to hearing impairment or reduced intelligibility. As we know, simultaneously planning a response whilst listening to the

incoming turn is cognitively demanding. When listening is challenging this further increases cognitive demands and reduces the resources available for timely response planning. We hypothesised that this would result in a change of strategy to late-planning, in order to reduce the strain on cognitive resources. Such a change in approach would have been demonstrated by a reduced difference between response times in the EARLY and LATE critical content conditions. However, our results suggest that listeners continue to plan early even when listening is challenging. One possibility is that whilst early planning is more cognitively demanding when listening is challenging, people with hearing loss can alleviate this demand by relying more on predictive contextual cues. Prior research suggests that in certain situations listeners with hearing loss show greater reliance on contextual cues compared to people with normal hearing (Benichov et al., 2012; Pichora-Fuller et al., 1995). It could be the case that, listeners use these compensatory cues to free up cognitive resources to support early planning. Alternatively, it could be that early planning is so ingrained that even when resources are depleted it is prioritised.

### ***Exploratory analysis of excluded trials***

A substantial proportion of trials were excluded because participants either did not respond within 2000ms after question offset (26.57%), because the answer was uninterpretable to the experimenter (0.35%), or because participants indicated they did not know the answer/ hear the question (0.27%). Our analysis showed exclusions were higher at the adaptive than the 60 dB presentation level for NH listeners and were higher for LATE than EARLY items at both presentation levels; at 60 dB, exclusions were also higher for HI than for NH listeners. In sum, there was an increase in removed trials when listening conditions were more challenging (for NH in the adaptive block, and for HI relative to NH at 60 dB), consistent with exclusion rates primarily reflecting

task difficulty. Importantly, we did not observe significant interactions (i.e., the difference in LATE/ EARLY trials being excluded was similar across groups and presentation levels), suggesting that exclusions did not disproportionately affect one condition.

It is also possible that some excluded trials could reflect very late response initiation (e.g., delayed planning). Because recordings ended at +2000ms, we cannot determine whether excluded trials would have contained valid responses after this window. However, response times were typically well within the recording window (median onsets  $\sim -89$  to 358ms for EARLY and  $\sim 535$  to 832ms for LATE across hearing groups/ presentation levels), making a widespread strategy shift toward very late responding unlikely.

### ***Strengths and limitations***

One limitation of the current design concerns the instruction for participants to respond, “as quickly as possible”, that is, as soon as they expected the question to end. While this instruction was intended to encourage early/ anticipatory response planning (and to discourage a strategy of waiting until the end of the question), it may have prompted participants to prioritise speed over natural conversational timing or strategy. In everyday conversation, people do not have the explicit goal of responding as fast as possible; rather, they balance promptness with social appropriateness (e.g., avoiding overlap with the speaker). It is therefore possible that our reported negative response times, and lack of adjustment of planning strategy, thus are somewhat specific to the task and instructions selected, and that planning may function differently when immersed in the constraints and goals inherent of natural conversation.

While speech was presented at 60 dB, our paradigm was able to replicate the

hearing-related delays that have been found in real conversation (Petersen et al., 2022; Sørensen et al., 2024), the paradigm likely prohibited many of the reasons for temporal adjustments to turn-taking that occur in natural conversation. In conversation, interlocutors produce their responses at specific times based on not only the information gained from the incoming turn but also based on the message they intend to convey and social etiquette. By asking participants to respond with the ‘correct’ answer to a simple and constrained question (rather than respond in a way to continue the conversation, build a relationship, or signal enthusiasm or misunderstanding), we likely removed many pragmatic reasons for temporal adjustments. Furthermore, participants may have become accustomed to the EARLY vs. LATE question types throughout the task, which do not reflect the nuance or flexibility of speech in real-life conversation, and may have created a bias towards early planning when it was possible. Nonetheless, the fact that both participant groups were able to engage in early planning to an equal degree, and that they both chose to do so when the information was available (even when listening was challenging), remains an important finding.

Finally, it is important to note that in conversation, participants have access to other cues signifying an upcoming turn end (head turns, eye-movements etc), and that these cues may be particularly salient for hearing impaired interlocutors. The removal of these cues in the question answering task may thus have impacted the HI group disproportionately, if they rely more on such cues to compensate for auditory degradation. While this could have increased their difficulty in predicting the end of the current turn and timing their responses in the 60 dB condition, we note that the two groups performed equally at in the adaptive threshold condition, reducing the risk that such strategy differences affected our findings more broadly.

## ***Conclusion***

In summary, we investigated how the timecourse of response planning is affected by challenging listening conditions. We showed that despite HI listeners being slower overall to respond to questions at 60 dB than NH listeners, both groups responded faster when the critical content appeared EARLY. Furthermore, we found no difference in response times between groups when listening at comparable levels of intelligibility (i.e., in the intelligibility-matched adaptive threshold condition), and no differences in the benefit of critical content occurring EARLY vs LATE, suggesting that delays relating to hearing loss may be explained by a simple reduction of intelligibility. As such, it appears that listeners plan early even when listening demands increase, and this strategy is unaffected by hearing status.

## References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Barthel, M., & Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science*, 43(7), <https://doi.org/10.1111/cogs.12768>
- Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, 7, <https://doi.org/10.3389/fpsyg.2016.01858>
- Benichov, J., Cox, L. C., Tun, P. A., & Wingfield, A. (2012). Word recognition within a linguistic context. *Ear & Hearing*, 33(2), 250–256. <https://doi.org/10.1097/AUD.0b013e31822f680f>
- Bögels, S., Casillas, M., & Levinson, S. C. (2018). Planning versus comprehension in turn-taking: Fast responders show reduced anticipatory processing of the question. *Neuropsychologia*, 109, 295–310. <https://doi.org/10.1016/j.neuropsychologia.2017.12.028>
- Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports*, 5(1), 12881. <https://doi.org/10.1038/srep12881>
- Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition*, 175, 77–95. <https://doi.org/10.1016/j.cognition.2018.01.015>

- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, 55(2), 230–240.  
<https://doi.org/10.1080/0163853X.2017.1330031>
- Corps, R. E., & Pickering, M. J. (2024). The role of answer content and length when preparing answers to questions. *Scientific Reports*, 14(1), 17110.  
<https://doi.org/10.1038/s41598-024-68253-6>
- Corps, R. E., Shehzad, M., McLaren, A. I., Fernandez, L. B., & Hadley, L. V. (2025). Effects of listening demand on turn-taking processes: Investigating turn-end prediction and verbal response planning. *Acta Psychologica*, 260, 105694.  
<https://doi.org/10.1016/j.actpsy.2025.105694>
- Fairs, A., Bögels, S., & Meyer, A. S. (2018). Dual-tasking with simple linguistic tasks: Evidence for serial processing. *Acta Psychologica*, 191, 131–148.  
<https://doi.org/10.1016/j.actpsy.2018.09.006>
- Fernandez, L. B., Pickering, M. J., Naylor, G., & Hadley, L. V. (2024). Uses of linguistic context in speech listening: Does acquired hearing loss lead to reduced engagement of prediction? *Ear & Hearing*, 45(5), 1107–1114.  
<https://doi.org/10.1097/AUD.0000000000001515>
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279. <https://doi.org/10.1111/1467-9280.00255>
- Huetting, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, 31(1), 80–93.  
<https://doi.org/10.1080/23273798.2015.1047459>
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.  
<https://doi.org/10.1016/j.cognition.2002.06.001>

- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.  
<https://doi.org/10.1080/01621459.1995.10476572>
- Kemper, S., Herman, R. E., & Lian, C. H. T. (2003). The costs of doing two things at once for young and older adults: Talking while walking, finger tapping, and ignoring speech or noise. *Psychology and Aging*, 18(2), 181–192.  
<https://doi.org/10.1037/0882-7974.18.2.181>
- Kubose, T. T., Bock, K., Dell, G. S., Garnsey, S. M., Kramer, A. F., & Mayhugh, J. (2006). The effects of speech production and speech comprehension on simulated driving performance. *Applied Cognitive Psychology*, 20(1), 43–63.  
<https://doi.org/10.1002/acp.1164>
- Lee, M. D., & Wagenmakers, E. J. (2014). Bayesian cognitive modeling. *Cambridge University Press*. <https://doi.org/10.1017/CBO9781139087759>
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6.  
<https://doi.org/10.3389/fpsyg.2015.00731>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477.  
<https://doi.org/10.1121/1.1912375>
- Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear & Hearing*, 39(2), 204–214.  
<https://doi.org/10.1097/AUD.0000000000000494>
- Petersen, E. B., MacDonald, E. N., & Sørensen, A. J. M. (2022). The effects of hearing-aid amplification and noise on conversational dynamics between normal-hearing and hearing-impaired talkers. *Trends in Hearing*, 26.  
<https://doi.org/10.1177/23312165221103340>
- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear & Hearing*, 37(1), 5S–27S. <https://doi.org/10.1097/AUD.0000000000000312>

- Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97(1), 593–608. <https://doi.org/10.1121/1.412282>
- Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö, Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7, <https://doi.org/10.3389/fnsys.2013.00031>
- Schad, D. J., Nicenboim, B., Bürkner, P.-C., Betancourt, M., & Vasissth, S. (2023). Workflow techniques for the robust use of Bayes factors. *Psychological Methods*, 28(6), 1404–1426. <https://doi.org/10.1037/met0000472>
- Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension – An fMRI study. *Cerebral Cortex*, 22(7), 1662–1670. <https://doi.org/10.1093/cercor/bhr249>
- Sjerps, M. J., & Meyer, A. S. (2015). Variation in dual-task performance reveals late initiation of speech planning in turn-taking. *Cognition*, 136, 304–324. <https://doi.org/10.1016/j.cognition.2014.10.008>
- Sørensen, A. J. M., Lunner, T., & MacDonald, E. N. (2024). Conversational dynamics in task dialogue between interlocutors with and without hearing impairment. *Trends in Hearing*, 28, <https://doi.org/10.1177/23312165241296073>
- Sørensen, A. J. M., MacDonald, E. N., & Lunner, T. (2020). Timing of turn taking between normal-hearing and hearing-impaired interlocutors. *International Symposium on Auditory and Audiological Research: Auditory Learning in Biological and Artificial Systems*, 37–44.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26), 10587–10592. <https://doi.org/10.1073/pnas.0903616106>
- Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale – Revised.
- Wendt, D., Kollmeier, B., & Brand, T. (2015). How hearing impairment affects sentence comprehension: Using eye fixations to investigate the duration of

speech processing. *Trends in Hearing*, 19,  
<https://doi.org/10.1177/2331216515584149>

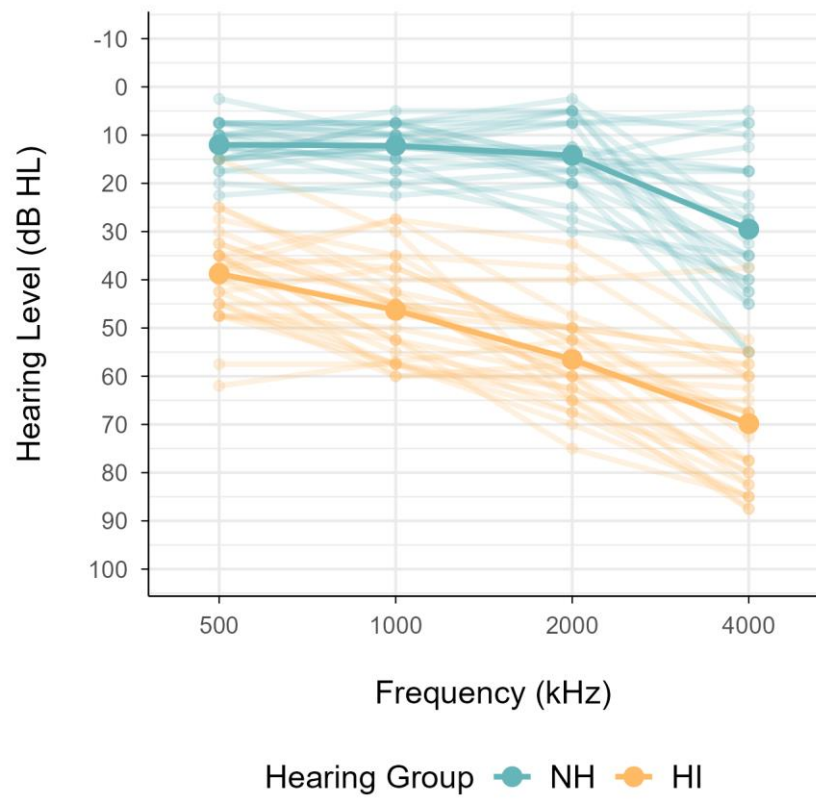
**Table 1.** Averages for participant demographics, intelligibility scores, adaptive presentation level, and processing speed.

|                                                             | <b>Hearing Group</b>  |                         |           |
|-------------------------------------------------------------|-----------------------|-------------------------|-----------|
|                                                             | <b>Normal Hearing</b> | <b>Hearing Impaired</b> |           |
| <b>Age</b>                                                  | 63.92 (6.68)          | 69.75(7.02)             | p = 0.33  |
| <b>Better Ear Average</b>                                   | 15.10 (4.79)          | 48.21 (7.35)            | p < 0.001 |
| <b>60 dB Intelligibility (% Words Correct)</b>              | 95.36 (8.57)          | 86.67 (4.71)            | P < 0.001 |
| <b>Adaptive Threshold Intelligibility (% Words Correct)</b> | 80.83 (5.00)          | 84.76 (7.16)            | p = 0.36  |
| <b>Adaptive Threshold Presentation Level (dB)</b>           | 29.20 (3.75)          | 55.82 (7.59)            | p < 001   |
| <b>Processing Speed (DSST)</b>                              | 45.68 (1.03)          | 41.86 (0.82)            | p = 0.16  |

**Table 2.** Percentage of excluded trials by listening condition and hearing group.

| <b>Presentation Level</b> | <b>Normal Hearing</b> | <b>Hearing Impaired</b> |
|---------------------------|-----------------------|-------------------------|
| <b>60 dB</b>              | 16.20 %               | 28.51 %                 |
| <b>Adaptive Threshold</b> | 29.67 %               | 33.45 %                 |

**Figure 1.** Pure-tone thresholds for NH and HI participants. Bold line represents the mean.



**Figure 2.** Average response times (ms) for normal hearing and hearing-impaired participants at 60 dB and adaptive presentation levels. Error bars represent standard error of the mean.

