Full Length Article

Continuous tracking of effort and confidence while listening to speech-in-noise in young and older adults

Chiara Valzolgher [a,*], Alessandra Federici [b], Elena Giovanelli [a], Elena Gessa [a], Davide Bottari [b], Francesco Pavani [a,c]

[a] Center for Mind/Brain Sciences – CIMeC, University of Trento, Rovereto, Italy
[b] MoMiLab, IMT School for Advanced Studies Lucca, Italy
[c] Centro Interuniversitario di Ricerca "Cognizione, Linguaggio e Sordità" – CIRCLeS, Trento, Italy

ABSTRACT

Reporting discomfort when noise affects listening experience suggests that listeners may be aware, at least to some extent, of adverse environmental conditions and their impact on listening experience. This involves monitoring internal states (effort and confidence). Here we quantified continuous self-report indices that track one's own internal states and investigated age-related differences in this ability. We instructed two groups of young and older adults to continuously report their confidence and effort while listening to stories in fluctuating noise. Using cross-correlation analyses between the time series of fluctuating noise and those of perceived effort or confidence, we showed that (1) participants modified their assessment of effort and confidence based on variations in the noise, with a 4 s lag; (2) there were no differences between the groups. These findings imply extending this method to other areas, expanding the definition of meta-cognition, and highlighting the value of this ability for older adults.

1. Introduction

Acoustic environments that surround us are often noisy. Multiple acoustic sources emit sounds and noises from various locations in space. What reaches the ears is a composite auditory mixture of sound streams that the brain must distinguish to comprehend the spoken sentence, the isolated words, or generally the sound of interest (Bregman, 1994; Sussman, 2017). The intricate process of segregating each sound source hinges on both the fidelity of the auditory input captured by the ears and the efficacy of higher-order cognitive skills, such as attention and working memory (Bronkhorst, 2015). Although people often manage to listen and engage in conversations in noisy environments, this does not imply that the task is achieved without significant effort. For instance, listening in a noisy environment becomes a demanding and effortful task when the auditory system is not optimally functional (e.g., for older adults, Anderson et al., 2013; Gosselin and Gagne, 2011), or when the noise reaches high levels of intensity (e.g., in a classroom full of students, Howard et al., 2010), or when visibility of lip movements is occluded (e.g., by face masks, Giovanelli et al., 2021).

Reporting discomfort when noise affects the listening experience suggests that listeners may be aware, at least to some extent, of the adverse environmental conditions and their impact on the listening experience. This concept extends beyond performance-based descriptions, such as word identification, to encompass their ability to monitor internal states, such as *perceived effort* and

---

confidence in what has been heard. The ability to track one's own internal states (i.e., metacognitive function; Efklides, 2008, Flavell, 1979) has often been investigated by assessing retrospective judgments. For instance, researchers adopted standardized questionnaires (Gatehouse & Noble, 2004) or self-reported questions with rating scales to measure feelings of effort or confidence associated with a cognitive task (but see also Gagne et al., 2017). Rabini and colleagues (2020) asked participants to evaluate their confidence on a Likert scale at the end of each trial while engaged in a sound localization task (Rabini et al., 2020). Similarly, Giovanelli and colleagues (2021) instructed participants to perform a hearing in noise task (i.e., they were asked to pay attention only to a male talker speaking simultaneously to other female distractors) and to evaluate at the end of each trial their confidence in what they heard and their listening effort, using a Likert scale from 0 to 9. These subjective reports are always collected at the end of an extended experiential segment (i.e., a trial or a block of the experiment).

While these indices can summarize confidence and perceived effort at the end of the extended experiential segment, they do not capture the continuous changes of these metacognitive dimensions that people could experience while listening in real-time. A noteworthy exception in this direction is the measure of pupil dilation. Pupil size increases during the execution of cognitively demanding tasks (Laeng & Alnaes, 2019; Wendt et al., 2017; Koelewijn et al., 2015; McGarrigle et al., 2021). However, the alteration of the pupil size does not necessarily reflect the conscious sensation of effort. It can be the result of a combination of factors, including attention, engagement, arousal, anxiety, as well as effort (Nunnally et al., 1967; Pichora-Fuller et al., 2016). In addition, pupil dilation is typically measured in response to short-lasting auditory events, typically ranging from 2 to 6 s (Winn et al., 2018; Saksida et al., 2022). Thus, monitoring of real-time changes is generally not easily addressed with this approach.

The first aim of the present study was to quantify, for the first time, continuous self-report indices that tracks one's own internal states. In particular, the feeling of effort and confidence experienced when listening to stories with noise fluctuating over time.

Listening in noisy conditions is particularly stressful for older adults, whose ability to hear in noise is often weakened or impaired, as in cases of age-related hearing loss (ARHL; Bowl & Dawson, 2019; Fostick & Schneider, 2022; Nadol, 1993). Older adults often report difficulties and frustrations when listening in noisy contexts (Pichora-Fuller, 1997; Glyde et al., 2011) with negative consequences for their quality of life and sociability (Mick, Kawachi & Lin, 2014). To date, little investigation has been done to determine if the worsening of hearing performance with age is also accompanied by poor ability to evaluate one's internal state while listening (see Giovanelli et al., 2023 for a recent exception). Most of the studies investigating the differences between young and older adults have focused on measures of effort, conceptualized as the "extra" mental processing resources required in adverse listening conditions (Degeest et al., 2015; see Lemke & Besser, 2016 for terms). They adopted a dual-task paradigm to measure effort during listening speech-in-noise and observed that older adults expend more listening effort than young adults (Degeest et al., 2015). Other research focused on the evaluation of internal states and tested the effect of age by asking participants to rate their listening confidence using classical Likert scales. For instance, Giovanelli and colleagues (2023) documented that younger adults showed higher confidence scores than older adults. However, to date, no studies have investigated the difference between young and older adults in continuous tracking of effort and confidence.

Taking advantage of measuring metacognitive states, our second aim was to investigate age-related differences in the capabilities of continuously tracking one's own internal states (i.e., effort and confidence) when listening to speech-in-noise. To this aim, we compared a group of young and a group of older adults with typical hearing. Deepening the effects of age on this ability is crucial to investigate whether an effective continuous tracking of the internal state should be an important resource for older people to face the challenging situations they may encounter in everyday life.

To pursue these two aims, we asked young and older participants to continuously report their confidence and effort while listening to stories embedded in fluctuating noise, that is, noise levels that changed throughout each story reaching different signal-to-noise ratios (SNRs). To extract the time series of the subjective reports, both groups were instructed to listen to two stories, each lasting approximately 3 min, and simultaneously to press one out of 5 keys (1, 2, 3, 4, 5) on the keyboard to continuously report their subjective feelings of confidence or effort. The time series of fluctuating noise were introduced to ensure that confidence and effort would fluctuate during listening. In addition, they provided the objective change in the environment against which we compared the subjective reports of each participant. Importantly, to control for potential age-related changes in tolerance to speech in noise perception the SNRs were individually adapted according to each participant's ability to hearing-in-noise, so that all participants could listen to the story with ease (at least 90 % intelligibility). This methodological choice aimed to avoid that the task proved too overwhelming for the participant – and particularly for the older adults – ensuring that participants could keep the focus of their attention on the internal states (confidence and effort) rather than comprehension of the story.

To characterise the relation between the two time-series (the objective change in the environment and the subjective report of the internal state) we used cross-correlation analysis (see paragraph 2.4. Analyses in Method for details). This approach allows us to capture (1) the overall degree of synchrony between the participants' responses and the environmental fluctuations (SNR changes); (2) any time delay of the internal state subjective tracking. With respect to our first question (can people track changes in their internal states of confidence and effort when listening in noise?), we hypothesized that the degree of synchronization between the subjective reports of these internal states and the fluctuations in SNR will vary across participants. Rather than expecting perfect, zero-lag synchronization or a complete lack of synchronization, appear both as unrealistic options. We anticipate, instead, that the relationship will fall within a range that reflects meaningful, albeit not perfect, alignment. Specifically, we predict that due to the known effects of background noise on both effort and confidence (e.g., Howard et al., 2010), participants will demonstrate a moderate level of synchronization, indicating their ability to continuously monitor and adjust their internal states in response to changes in the acoustic environment.

With respect to our second question (do older adults differ from young adults in tracking their internal states?) there are currently not many studies comparing the metacognitive skills of young versus older adults in relation to listening in noise (with the exception of

Giovanelli et al., 2023, mentioned earlier). However, some research has noted that older adults are less able to detect errors and changes in their cognitive processes. For example, studies investigating metacognition in the field of memory (Tullis & Benjamin, 2012; Hertzog et al., 2002) and error awareness in a go/no-go response inhibition paradigm (Sim, Brown, O'Connell, & Hester, 2020) have shown this decline. These studies hint at possible differences that could emerge using continuous measures, which may detect age differences not seen in studies measuring metacognition of listening in noise as an interaction between performance and confidence (Giovanelli et al., 2023). Specifically, the two parameters we extracted from cross-correlation analysis permit us not only to determine whether older adults can perform metacognitive tracking but also to identify any differences in terms of the level of synchrony and latency with which they do so compared to younger adults. On the one hand, the overall degree of synchrony could help us to detect any differences in the strength of correlation (reflecting higher monitoring abilities), which may be more sensitive to possible deterioration due to aging that has not yet been documented by previous studies on listening in noise. Measuring time delay, on the other hand, could permit to reveal any potential timing weakness that older adults might experience compared to younger individuals in monitoring their internal sensations.

2. Material and methods

2.1. Participants

Sixty-four participants took part in the study, divided into two age groups: young (N=32; mean age = 22.9 years, sd = 2.8, range = [20–32]; 10 males, 22 females) and older adults (N=32; mean age = 67.8 years, sd = 5.8, range = [61–82]; 7 males, 25 females). There were no differences in terms of gender distribution between the two groups ($X^2 = 0.72$, $p = 0.40$). There were 5 left-handed attendees (4 in the young group). The ages of the two groups differ statistically, as revealed by running an independent sample t-test (Welch, W (45) = 39.58, $p < 0.001$). There are no previous studies investigating continuous subjective ratings as a function of age; hence, we could not estimate effect size. Thus, we determined the sample size based on the previous study comparing young and older adults on metacognitive appreciation of listening speech-in-noise (Giovanelli et al., 2023).

All methods were performed following the Declaration of Helsinki (1964, amended in 2013), and participants provided their informed consent. To exclude hearing deficits, pure tone audiometry (PTA) was measured using an audiometer (Graison Stadler GSI 17 Audiometer) at different frequencies (250, 500, 1000, 2000, and 4000 Hz) tested separately for each ear. Participants with PTA below or equal to 25 dB/HL were considered to have normal hearing (Nadol et al., 1993). As expected, the average PTA was higher for older adults (mean = 13.93 dB HL, sd = 4.9) compared to young participants (mean = 5.6 dB HL, sd = 4.2; independent sample t-test, $t(62)$ = 7.24, $p < 0.001$). Older adults also completed the Italian version of the Montreal Cognitive Assessment (MoCA) to exclude any cognitive impairment associated with aging. All older adults reported no cognitive deficits (N=32; mean = 28.2, sd = 1.3, cut-off = 26, scored with the correction of Santangelo et al., 2015).

2.2. Stimuli

Preliminary threshold assessment. Speech stimuli consisted of a set of 50 words from the Italian version of the Matrix Test, used
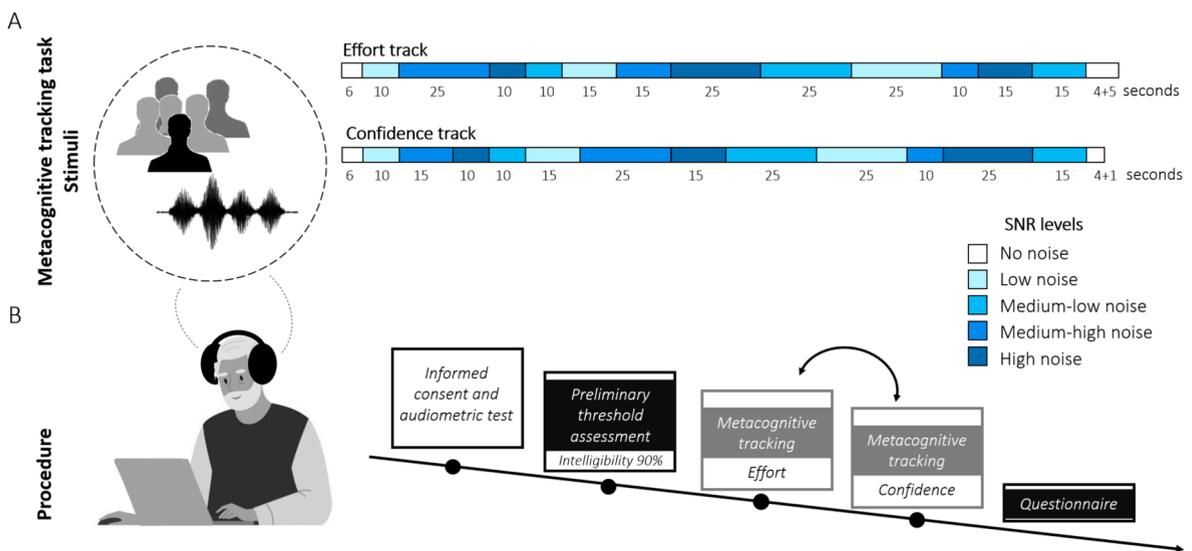

Fig. 1. Stimuli and experimental timeline. (A) The stories were embedded in multi-talker babble noise; the structure of the track for each judgment type (i.e., effort and confidence of the metacognitive tracking task is shown with the different SNR levels and the length of each interval). (B) Schematic description of the experimental timeline (the order of effort and confidence judgment was randomized).

in audiological practice to assess hearing-in-noise abilities (Puglisi et al., 2015). The words of the Matrix Test are combined into phrases with correct grammatical and syntactical structure but without semantic predictability to avoid the effects of semantic competence on recognition. An example of a possible phrase is "Sofia sends two useful boxes" (translation of the Italian original sentence: "Sofia manda due scatole utili").

Metacognitive tracking task. Continuous speech stimuli were two Italian short stories, lasting approximately 3 min each, selected from an Italian book of short novels ("I Tacchini non ringraziano" by Andrea Camilleri; selected stories: "L'incantesimo della volpe" and "Il cardellino"). One story was used for effort evaluation and the other for confidence evaluation, randomized and balanced across participants. Crucially, each story was embedded in noise. As illustrated in Fig. 1A, four signal-to-noise ratio (SNR) levels were used: high noise, medium–high noise, medium–low noise, and low noise (they were coloured according to the SNR on a blue tone scale). Each SNR level was created for each individual based on the performance in the preliminary threshold assessment (see below). Note that, for each SNR level, the speech intensity consistently remained higher than the threshold at which each participant correctly identified 90 % of the words. Each noise level was presented for a total of 50 s divided into three intervals of 25, 15, and 10 consecutive seconds. As a consequence, each track comprised 12 different intervals of speech-in-noise (4 SNR x 3 segment length). Note that the order of the SNR sequences was identical for the two tracks, but we changed the sequence on the basis of segment length. To make the flow of the acoustic track smooth and more natural, each interval had a 3-second linear ramp that gradually brought the SNR level from that of the previous interval to that of the desired one. Note that we generated two sets of interval sequences: one for assessing effort and the other for confidence. These sequences maintained their association with the respective judgment types (i.e., confidence and effort) while the two stories were randomized. To allow participants to easily recognize the voice telling the story and better track the target speech, each audio stream starts without noise (6 s). Similarly, each audio stream ends with a few seconds without noise (5–9 s).

The noise employed in both tasks (i.e., Preliminary threshold assessment and Metacognitive tracking task) was multi-talker babble noise obtained through the combination of five different voices (two females, three males) reading random passages from a fiction book (*La Strada*, McCarthy, 2006/2014). Each voice was recorded in a single file, and the files were then superimposed to obtain the multi-talker babble noise (Bednaya et al., 2022).

The speaker of the 50 words and the two stories was always the same Italian woman whose dictation was trained. The recording was performed in a sound-attenuated chamber (BOXY, B-Beng s.r.l., Italy) at the IMT School for Advanced Study Lucca and it was registered using an iPhone 7 (camera with 12MP, video resolution in HD, 720p with 30 fps, at a sampling frequency of 48000 kHz) and an external condensation microphone (YC-LM10 II, Yichuang). For the stories' stimuli, the speaker waited a few seconds, read the title, and then the whole story. Recordings were imported in iMovie (version 10.3.1), and 100 % noise reduction was applied. For each story, we cut the file to have 2 s of silence before the title of the story and approximately 5–9 s of silence at the end of the story. Then, the audio was imported into Audacity® (version 2.4.2, https://www.audacityteam.org/ using FFmpeg and lame functions to isolate the audio from the video) and pre-processed to have a mono audio file set at 32-bit sample. Finally, we performed RMS equalization to achieve an equal loudness for all the stimuli within each task.

2.3. Apparatus

The experiment was presented using a DELL G5 15 laptop (Processor Intel Core i7) and over-ear headphones (Sennheiser HD 650 S; HiFi, frequency range: 10–41.000 Hz). Both the preliminary threshold assessment and the metacognitive tracking task were delivered through scripts coded in Matlab (ver 2021b).

2.4. Procedure

The portability of our virtual audio-visual setup allowed testing of participants either in the laboratory or at their homes, always ensuring the task was administered in a quiet environment. Participants started the experimental session by giving informed consent. Recall that, before starting the experiment, we measured the pure tone audiometry of each participant to exclude possible hearing deficits. Furthermore, older adults completed the MoCA to exclude any cognitive impairment associate with ageing. Then, they performed the hearing-in-noise task (i.e., Preliminary threshold assessment), the two metacognitive tracking tasks (confidence and effort judgment), and finally, they completed a questionnaire aimed at assessing participants' comprehension of the stories (Fig. 1B).

Preliminary threshold assessment. Participants sat in front of the computer. They were asked to listen to 20 sentences embedded in noise and to select from a grid displayed on the screen the correct sentences by using the mouse (note that the experimenter could move the mouse when older adults were not familiar with its use). Thus, the staircase procedure aimed at fixing the individual SNR so that each participant performed the metacognitive tracking task at a comparable difficulty level. This procedure was implemented through an algorithm that builds upon the formula of the adaptive procedure described in Brand & Kollmeier (2002) and fixed the target performance level at 90 % correctly reported words. Each block comprised 20 phrases of 5 words presented in informational noise (multi-talker babble noise), and the audio level was adjusted by modifying only the volume of the target voice, starting from SNR set at 0 (i.e., noise and target at the same volume). The first procedure was considered familiarization, while the second one was used to extract the SNR. The rationale of calculating the SNR at which participants yield 90 % correct is to use it as a base to create the different levels of noise (high noise = 90 % + 6 dB, medium–high noise = 90 % + 12 dB, medium–low noise = 90 % + 18 dB, and low noise = 90 % + 24 dB) presented during the stories. We decided to maintain at least 90 % intelligibility throughout the task to be sure that the stories delivered in the metacognitive tracking task were audible for each participant in a similar way and avoid different acoustic experience difficulty between participants. This approach aimed to help participants focus on their feelings of effort and confidence rather than being influenced by noticeable fluctuations in intelligibility (i.e., whether or not they could hear the speech clearly).

Additionally, since there are age-related changes in tolerance to speech in noise perception, we aimed to keep the task at a level of intelligibility tolerable for both groups. This was to ensure that fatigue, which could emerge in the older group due to prolonged exposure to a complex and poorly tolerated noisy listening environment, did not create a confound influencing our metacognitive measures.

Metacognitive tracking task. Participants were asked to listen to two different stories in succession. During one story, they were instructed to focus on their perceived effort and rate it using a Likert scale ranging from 1 (low effort) to 5 (high effort). During the other story, they were asked to focus on their confidence and rate it using a Likert scale ranging from 1 (low confidence) to 5 (high confidence). The order of the stories and the focus on either effort or confidence were randomized across participants. Throughout both stories, participants were required to press one of five keys (1, 2, 3, 4, 5) to report their target feeling and continually monitor it. At the beginning of each story, they were instructed to press 1 when rating effort and 5 when rating confidence. The logic was to start the evaluation from the same value for each participant (i.e. baseline) and instruct them to modify it according to any changes in their feelings. The stories began without any background noise, and participants were asked to change the key they were pressing whenever they noticed a change in their listening effort or confidence. They were explicitly told to concentrate on their internal states and not to change the button if they did not perceive any changes, without explicitly referring to SNR changes. Effort refers to the sensation of needing to exert more energy to complete the task, while confidence pertains to the assurance that what they heard was exactly what was spoken, focusing on the perceptual experience rather than the meaning of the sentences.

Questionnaire. After the metacognitive tracking task (i.e. evaluating both stories), we assessed participants' comprehension of the stories using an 8-question questionnaire for each story, employing a four-alternative forced-choice task. Each question had a correct answer, a wrong answer semantically related to the correct one, a wrong answer phonetically related to the correct one, and an incongruent answer. The order of presentation of the four answers was randomized. Note that both groups were not informed about the questionnaire concerning the stories' content before listening to them. The instruction for the metacognitive task was to focus on the internal feeling of effort or confidence and to continuously report it, and participants were not informed that they were asked about the stories' contents.

2.5. Analyses

To measure the impact of the adverse listening conditions and the subjective tracking, we analyzed, at the individual level, the similarity between noise fluctuations occurring during the story presentation and the metacognitive tracking of each participant, continuously reporting listening effort or confidence. In order to estimate the magnitude of this similarity as a function of lag (see Barajas et al., 2021), we measured the cross-correlation between the two time-series (i.e., noise fluctuation and subjective continuous report). We extracted two indices for each participant: the maximum value of the cross-correlation (r) and its associated time lag (reflecting the time point in which the noise fluctuation and metacognitive tracking were most correlated). Note that we normalized the noise fluctuation sequence to obtain r values between 0 and 1. Moreover, only for the confidence index we reversed participants' responses by using the formula 6 minus the value, to have the same range of values as the noise fluctuations in which higher values reflect a higher level of uncertainty (see Fig. 2D). To validate the significance of the association between the noise fluctuation and subjective continuous reports (i.e., effort or confidence), we contrasted the indexes obtained from each cross-correlation with the respective permutated indexes. We calculated the permutated indexes for each participant by correlating the noise fluctuation and a permutated vector of their subjective continuous report. That is, for each participant, we performed a cross-correlation between the noise fluctuation and a permutated version of each continuous report (i.e., effort or confidence). Then, we contrasted the magnitude and the time lag of the real cross-correlation with the magnitude and the time lag of its permutated cross-correlation version. The comparison was performed using a paired-sample test (Wilcoxon, as the variables are not normally distributed).

To describe the participants' ratings during the metacognitive tracking task, we also extracted: (1) the average values of the response provided during each story as an index of overall effort or confidence required by the task; (2) the peak/extremity of confidence and effort reached during the task, which corresponds to the minimum value (when considering confidence; i.e., peak uncertainty corresponds to the minimum value) or the maximum value (when considering effort; i.e., peak effort corresponds to the maximum value); (3) the number of answer changes, i.e., the number of times they changed their answer during the stories, as an index of the tendency to frequently adjust one's answers; and (4) the changes adaptation time (s), i.e., the average of the time to adjust the subjective judgment when an SNR changes. To calculate the latter index, we did not consider the first 3 s in which the noise increased to reach the target SNR. The rationale was to have a measure of the reaction to SNR changes; thus, we preferred to calculate it when the ramp was finished, and the SNR for that segment was reached and stable.

Parametric and non-parametric tests were used for statistical analyses, which were run using JASP 0.17.2.1. For all analyses, statistical significance was set to $p < 0.05$ (two tails). Plots were generated by using R (version 1.0.143). Note that we removed 1 participant (id 13, older adult) who did not change their response across time when asked to judge either effort or confidence. We decided to remove this participant as their data matched the most extreme scenario (no changes detected), which was unexpected and our method presupposes a minimum variability of change. This does not mean that cross-correlation technique cannot extract numerical indices even in the absence of variation in one of the traces. However, in this context, doing so would yield indices that have no relevance to our experimental question.

3. Results

Older adults needed a more favourable SNR (mean = -0.07, sd = 0.85) to achieve 90 % accuracy in the preliminary threshold
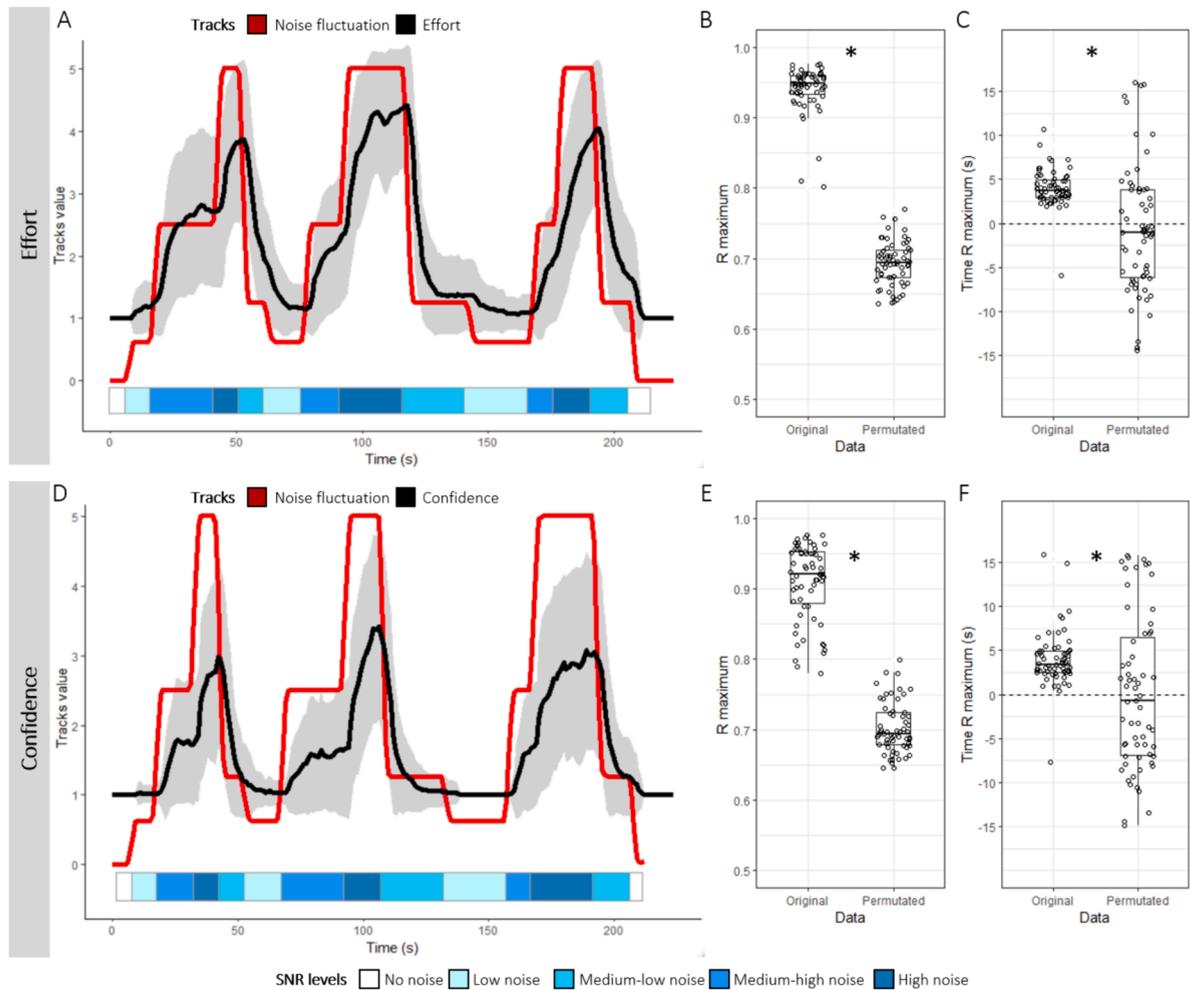
Fig. 2. Perceived effort and confidence during the metacognitive tracking task. Plots show data aggregated across both age groups. A and D) the average trace of the perceived effort (A) and confidence (D) reported by the participants as a function of time (s) (black) and the numerical parameter adapted to a scale ranging from 0 to 10, which regulated the modulation of the noise intensity (red). The shadows represent the standard deviation. Note that to represent noise fluctuation (in red), we normalized volume fluctuation to range from 0 to 5 in order to represent it in the graph. The noise intervals are represented in different colours on the background. B and E) Maximum values of R were obtained by running a cross-correlation analysis, considering the noise track and the effort (B) or confidence (E) tracks as a function of the data type (original or permutated). C and F) Time lag of the Maximum R values obtained by running a cross-correlation analysis considering the noise track and the effort (C) or confidence (F) tracks as a function of the data type (original or permutated). Asterisks indicate significant differences.

assessment, compared to young adults (mean = -1.25, sd = 1.29, $W(53.64) = 4.3$, $p < 0.001$). In addition, we documented a significant correlation between pure tone audiometry (PTA) and preliminary threshold assessment (*Spearman's rho* = 0.51, $p < 0.001$), meaning that the higher the PTA (i.e., worse hearing detection), the more positive the SNR in the preliminary threshold assessment. Both these results were largely expected and suggest that adapting noise changes to the actual capabilities of each participant was a correct methodological practice that permitted the control of individual differences, which could have compromised the comparison between these two populations in this task.

3.1. Metacognitive tracking

The first aim of the present study was to quantify the feeling of effort and confidence experienced when listening to stories in contexts with noise fluctuating over time. The average track of the perceived effort reported by participants as a function of time (s) is shown in Fig. 2A (black line), together with fluctuation in the volume of background noise (red line). The two tracks – i.e., noise fluctuation and subjective continuous report – are related, suggesting that participants were able to adjust their assessment of perceived effort based on the variations in background noise while listening to the story (Fig. 2A). To test this observation, we contrasted both the R maximum (Fig. 2B and 2E) and Time R maximum (Fig. 2C and 2F) with the permutated version by using a non-parametric paired-sample test (Wilcoxon, as the variables are not normally distributed, see Analysis for details). Testing the

difference between indices of the real effort cross-correlation and its permutated version, we observed that both the magnitude and the time lag differed significantly (R maximum: $t(62) = 35.27$, $p < 0.001$, Cohen's d = 4.44; Time R maximum: $t(62) = 4.96$, $p < 0.001$, Cohen's d = 0.63[1]). The same analysis was run by considering the confidence reported by each participant. Again, as visible in Fig. 2D, the trace of perceived confidence (note that we reversed participants' responses so that a higher value of the track corresponded to higher uncertainty) and of the noises fluctuated together, suggesting that participants were able to modify their assessment of perceived confidence based on the variation in the background noise while listening to the story. We observed significant differences both in the magnitude and time lag of cross-correlation (R maximum: $t(62) = 19.95$, $p < 0.001$, Cohen's d = 2.51; Time R maximum: $t(62) = 3.44$, $p < 0.001$, Cohen's d = 0.43[1]). Taken together, these results clearly indicate that the everchanging adverse listening conditions, SNR changes, resulted in changes in subjective indexes of listening effort and confidence and that participants were consistently (not randomly) reporting them.

3.2. Comparison between young and older adults

The second aim of the present study was to investigate age-related differences in the capabilities of continuously tracking one's own internal states (i.e., effort and confidence) when listening-in-noise. The average tracks of the perceived effort and confidence reported by the two groups (young in black and older adults in grey) as a function of time (s) are shown in Fig. 3A and B, together with fluctuation in the volume of the background noise (red line). To assess to what extent ageing adults can achieve metacognitive tracking of their listening effort and listening confidence comparable to those of younger participants, we proceeded by running independent sample *t*-test comparisons between the indices describing metacognitive tracking in older adults and young adults. The analyses conducted indicate that there are no significant differences between the older and younger populations in terms of the maximum value of R and the time lag for maximum values of R (as depicted in Fig. 3C-F, where the data for younger individuals are represented in black and for older adults in grey). Further examination of other parameters revealed that, except for the average response effort, there are no significant differences between the older and younger populations. We summarized the result in Table 1 by reporting for each index, the mean (and the standard deviation) values for each group, and the statistics. To further examine the likelihood of the measured data under the null vs. alternative hypotheses, we also ran Bayesian independent sample t-tests (see Bayesian Factor, BF, together with the corresponding level of evidence reported in Table 1. See Supplementary materials for other indices representations.).

Please note that if we were to correct the *t*-test for average effort for multiple comparisons using the Bonferroni method, the p-value cut-off would be 0.01 (0.05/5), rendering the *t*-test statistically non-significant.

Questionnaire. Results revealed that participants hardly remember the stories. They answered correctly at 54.6 % of the questions presented in the questionnaire (story listened while judging the effort: 50.4 %; story listened while judging the confidence: 58.8 %). No differences between the two age groups emerged. When considering the stories in which participants were asked to rate their effort, older adults responded correctly at 3.97 (sd = 2.02) out of 8 questions, while young at 4.10 (sd = 1.63) out of 8 questions ($t(62) = 0.27$, $p = 0.79$, Cohen's d = 0.07). When considering the stories in which participants were asked to rate their confidence, older adults responded correctly at 4.38 (sd = 1.54) out of 8 questions, while young at 5.03 (sd = 1.73) out of 8 questions ($t(62) = 1.60$, $p = 0.11$, Cohen's d = 0.40).

4. Discussion

The primary aim of the present study was to measure to what extent participants can continuously monitor their own internal states – perceived effort and confidence – experienced when listening to stories in noise. Our findings showed that all participants modified their assessment of perceived effort or confidence in real time based on the variations in the background noise while listening to the story. From now on, we will refer to this skill as metacognitive tracking. The second aim of the present study was to assess to what extent ageing adults can achieve metacognitive tracking of their perceived listening effort and listening confidence comparable to those of younger participants. We observed that there are no differences in this tracking ability between the two groups.

4.1. Metacognitive tracking ability: Exploring the online, continuous dimension of the individual's capacity to assess effort and confidence

In addition to increasing the methodological solutions available to describe the ability to track one's own internal state, our findings contributed to enriching the terminology describing metacognitive abilities. Metacognition has been defined by Flavell and colleagues (1979) as the high-level cognitive functions of thinking about one's cognitive processes and abilities (Palmer, David, & Fleming, 2014). It originally referred to two components: metacognitive *knowledge*, which indicates the general crystallized knowledge and beliefs about our abilities and cognitive functioning, and metacognitive *monitoring*, which indicates the ability to monitor and be aware of the outcome of ongoing cognitive processes (Ackerman & Thompson, 2017). Metacognition, from this latter standpoint, is defined as the ability to align our self-judgments to the actuality of our cognitive or physical performances (see also Fleming, 2024). While specifically related to the concept of metacognitive monitoring, the ability examined in this work does not necessitate the consideration (and consequently measurement) of performance indices. In our task, the focus lies on monitoring one's internal states

---

[1] Effort: We run the same analysis by eliminating the negative values in the Time R maximum (ID = 65 (older adults)) and we found Time R maximum: $t(61) = 4.99$, $p < 0.001$, Cohen's d = 0.63. Confidence: We run the same analysis by eliminating the negative values in the Time R maximum (ID = 57 (older adults)) and we found Time R maximum: $t(61) = 4.10$, $p < 0.001$, Cohen's d = 0.52.
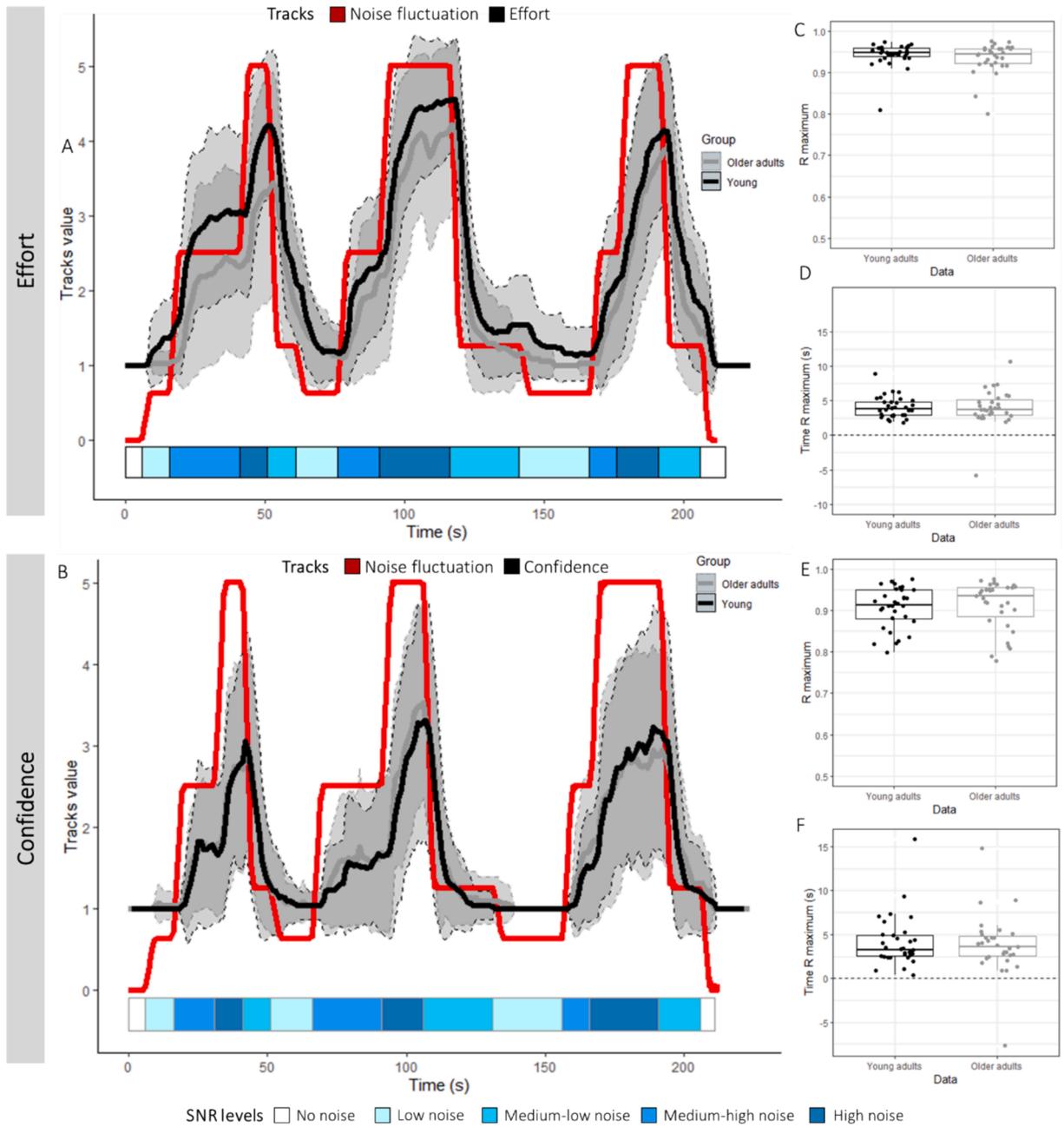
Fig. 3. Perceived effort and confidence during the metacognitive tracking task as a function of group. A and B) The average trace of the perceived effort (left) and confidence (right) reported by the participants as a function of time (s) and the numerical parameter adapted to a scale ranging from 0 to 10 which regulated the modulation of the track volume (red). These intervals are represented in different colours on the background. Older adults' tracks were represented in grey, while young adults' tracks in black. The shadows represent the standard deviation. C, D, E and F) Indices of metacognitive tracking task as a function of group (young, in black, and older adults, in grey). The two up graphs show the indices describing the effort rating, while the two bottom graphs show the indices describing the confidence rating. C and E) Maximum values of R obtained by running a cross-correlation analysis; D and F) Time lag of the Maximum values of R obtained by running a cross-correlation analysis. No significant differences emerged.

over time, and proficiency in such monitoring is evidenced by the ability to discern the implications of various noise fluctuations on these sensations. Therefore, this study introduced a new way to express metacognition on a continuous dimension: the metacognitive tracking ability.

To the best of our knowledge, this study represents the first attempt to explore this online, continuous dimension of the individual's capacity to assess one's own internal states (i.e., effort and confidence). Traditionally, paradigms addressing this aspect typically

Table 1

Metacognitive tracking indexes, mean (and the standard deviation) values for each group, and independent sample *t*-test.

Effort				
Indices	Mean (standard deviation)		Statistics	
	Older adults (N=31)	Young Adults (N=32)	*t*-test	BF^{10} / Evidence
R Maximum	0.94 (0.04)	0.94 (0.03)	$t(61) = 1.01$, $p = 0.32$ Cohen's d = 0.25	$BF^{10} = 0.38$ No evidence
Time R Maximum (s)[1]	4.02 (2.64)	4.02 (1.53)	$t(61) = 0.05$, $p = 0.96$ Cohen's d = 0.25	$BF^{01} = 0.26$ No evidence
Average	2.03 (0.59)	2.38 (0.60)	$t(61) = 2.35$, $p = 0.02$ Cohen's d = 0.59	$BF^{10} = 2.50$ Anecdotal
Standard Deviation	1.17 (0.42)	1.26 (0.28)	$W(51.51) = 1.03$, $p = 0.18$ Cohen's d = 0.26	$BF^{10} = 0.41$ No evidence
Maximum	4.45 (0.81)	4.66 (0.60)	$t(61) = 1.14$, $p = 0.26$ Cohen's d = 0.38	$BF^{10} = 0.45$ No evidence
Number of changes in the response	20.55 (7.05)	23.57 (8.47)	$t(61) = 1.63$, $p = 0.11$ Cohen's d = 0.41	$BF^{10} = 0.78$ No evidence
Changes adaptation time (s)	3.20 (1.32)	2.99 (1.46)	$t(61) = 0.62$, $p = 0.54$ Cohen's d = 0.16	$BF^{10} = 0.30$ No evidence
Confidence				
R Maximum	0.91 (0.06)	0.91 (0.05)	$t(61) = 0.43$, $p = 0.67$ Cohen's d = 0.11	$BF^{10} = 0.28$ No evidence
Time R Maximum (s)[2]	3.78 (3.47)	4.14 (2.93)	$t(61) = 0.45$, $p = 0.66$ Cohen's d = 0.11	$BF^{10} = 0.28$ No evidence
Average	4.37 (0.49)	4.32 (0.47)	$t(61) = 0.41$, $p = 0.68$ Cohen's d = 0.10	$BF^{01} = 0.28$ No evidence
Standard Deviation	0.87 (0.48)	0.92 (0.37)	$t(61) = 0.46$, $p = 0.65$ Cohen's d = 0.12	$BF^{01} = 0.28$ No evidence
Minimum	2.10 (1.25)	1.91 (0.96)	$t(61) = 0.68$, $p = 0.50$ Cohen's d = 0.17	$BF^{01} = 0.31$ No evidence
Number of changes in the response	20.00 (12.13)	22.75 (11.82)	$t(61) = 0.91$, $p = 0.37$ Cohen's d = 0.23	$BF^{01} = 0.37$ No evidence
Changes adaptation time (s)	3.65 (1.81)	3.59 (1.53)	$t(61) = 0.14$, $p = 0.89$ Cohen's d = 0.04	$BF^{01} = 0.26$ No evidence

[1]Effort: We run the same analysis by eliminating the negative values in the Time R maximum (ID = 65 (older adults)) and we found Time R maximum: t (60) = 0.69, p =0.50, Cohen's d = 0.17, BF = 0.57.

Confidence: We run the same analysis by eliminating the negative values in the Time R maximum (ID = 57 (older adults)) and we found Time R maximum: $t(60) = 0.03$, $p = 0.98$, Cohen's d = 0.008, BF = 0.26.

prompt participants to assess these perceptions at the end of the experience or after specific experiential segments (i.e., trials or blocks e.g., Rabini et al., 2020; Giovanelli et al., 2021, 2023, Visentin et al., 2023; Ohlenforst et al., 2017). While physiological indices such as pupil dilation can capture the continuous online aspect of listening effort (Laeng & Alnaes, 2019; Wendt, Hietkamp & Lunner, 2017), they represent indirect physiological measures that may not necessarily mirror the human awareness to discern the continuous oscillation of one's own internal states and feelings over time during the perceptual experience. Our study, by adopting an innovative methodological approach, outlined the ability of both young and older adults to adjust their self-ratings in response to variations in ambient noise while engaged in a listening task. This adaptive ability aligns with the theoretical framework of metacognition, specifically, the aptitude to monitor one's cognitive states.

From a certain perspective, measuring metacognitive tracking ability unveils a process that typically operates automatically and covertly. As anticipated in the previous paragraph, in typical experimental paradigms, participants were asked to rate their effort or confidence regarding a certain performance after specific experiential segments. Results indicated their proficiency in carrying out this task, demonstrating sensitivity to various conditions. Implementing this retrospective rating subtends the ability to recall the feelings experienced during a specific experiential segment. The cognitive system has, in a way, kept track of the effort and confidence experienced and, therefore, monitored them at that moment, engaging in a hidden metacognitive monitoring process. This process, in our experimental paradigm, becomes explicit as participants are asked to direct their attention entirely to their fluctuating internal states. Consequently, this task may be considered somewhat unnatural, as the typical monitoring of these states usually remains hidden during cognitive performances. However, it represents the most direct expression of one's feeling of effort and confidence experienced online during the task.

The unnatural request to focus 100 % on metacognitive tracking of effort and confidence had consequences in terms of performance. Participants struggled to remember the contents of the story, at least from what can be learned from the results obtained in the questionnaire proposed at the end of the experimental procedure. Both groups were only able to answer half of the questions correctly. The observed modest performance in the questionnaire should derive from the instruction provided by the participants: they were not explicitly invited to memorize the content of the stories in detail (i.e., we tested incidental memory). For this reason, the superficial listening of the contents by the participants is not surprising. Yet, to overcome this limit, future studies could propose new experimental paradigms. For example, they might ask participants to perform a double task, such as repeating or paying attention to the

content of the stories that they are listening to and simultaneously reporting their internal states. Since this is the first time that metacognitive tracking was measured, this stratagem was not chosen because we wanted to control for any confounds related to possible differences in the amount of resources used in the tasks between young and older adults.

The new direct continuous measure of subjective effort and confidence proposed in our study differs from other less direct indices adopted in previous experimental contexts. For instance, measuring changes in pupil size, acknowledged as an objective effort index, derives its objectivity from being a physiological indicator. Nonetheless, it is not a direct index as it encompasses a complex array of factors extending beyond listening effort. In this context, our subjective index can be viewed as a complementary measure for tracking the subjective continuous experience of effort and confidence over time. Given the complementarity of these indices, future studies could combine both the here-presented subjective direct measure of internal state with objective but indirect physiological signals such as pupil size. However, it should be noted that, as of now, pupillometry has not been adopted with long-duration stimuli such as stories. Therefore, future studies should consider a set of methodological precautions essential for obtaining reliable pupillometric data (see for notable exception studying emotional arousal Kaakinen & Simola, 2020). An example of combining indices can be found in the work of Visentin and colleagues (2023), who integrated reaction time, self-rating, and pupil indices in a single study, demonstrating that the sensitivity of these three measures to changes in signal-to-noise ratio (SNR) differs.

A physiological metric that has already been employed with long-term stimulation is instead brain oscillatory activity. Particularly, we referred to neural entrainment, which is defined as the alignment of the brain oscillatory activity to continuous stimulation (Lakatos et al., 2019). Neural entrainment to speech stimuli was classically investigated assessing the neural response related to the speech envelope (Luo & Poeppel, 2007; Obleser & Kayser, 2019) but also to linguistic features, such as phonemes and semantics (e.g., Di Liberto et al., 2015; Klimovich-Gray et al., 2023). In our view, this measure could be employed in combination with the continuous direct measurement of subjective effort and confidence that we have validated in this study. Incorporating different indices that reflect various aspects of cognitive and metacognitive processes within the same experimental paradigm may allow for the identification of the most sensitive measure under different experimental conditions.

4.2. Metacognitive tracking ability in young and older adults

Concerning the second objective of the current study, we found that the comparison between young and older adults did not yield any significant differences in the *meta*-cognitive tracking ability. To date, existing literature has predominantly focused on the impact of age on metacognition within the domain of memory. Numerous studies have found a decline in metacognitive ability associated with memory functions (for a review, see Hertzog et al., 2002). However, in the domain of acoustic perception, metacognitive abilities have remained relatively underexplored. A recent exception is the research conducted by Giovanelli and colleagues (2023), which explored metacognitive abilities in a speech-in-noise task. The authors measured metacognitive monitoring by extracting an index reflecting the coherence between participants' actual performance and their confidence in that performance (for details about indices see also Fleming & Lau, 2014). In this context, a proficient "metacognitor" is characterized by heightened confidence during proficient performance and diminished confidence during suboptimal performance. Furthermore, by proposing self-report questionnaires, they investigated metacognitive knowledge abilities. Notably, their findings failed to reveal any decline in metacognition in older adults compared to young adults, with certain results even suggesting superior metacognitive abilities in older adults.

Our study, in addition to Giovanelli's, examines a further facet of metacognitive abilities that has not been investigated before: the metacognitive tracking ability. This skill represents an additional aspect of self-awareness regarding one's abilities, requiring continuous and online monitoring of internal states. While investigating metacognitive aspects related to listening-in-noise, we might have expected results similar to those of Giovanelli and colleagues' study. However, this new dimension could also contribute to deepening any differences within this population. For instance, we expected that the documented cognitive slowing in the older population might influence the ability to promptly adjust one's judgment to SNR changes, even if they were able to detect them (Birren & Fisher, 1992). Our findings align with previous observations, indicating no significant differences between young and older adults in terms of their metacognitive abilities related to hearing in noise. Specifically, neither the R maximum nor the Time of the R maximum exhibited disparities between the two age groups, suggesting that the capacity to adjust internal state ratings as a function of background noise fluctuation remains stable with aging. Furthermore, our analysis did not reveal any slowing down in metacognitive tracking of older adults as compared to the young. To sum up, the scenario that emerged matched with our hypothesis to observe a moderate level of synchronization, indicating their ability to continuously monitor and adjust their internal states in response to changes in the acoustic environment. Specifically, we observed that for both groups, the greatest synchronization occurred with a time delay of 4 s. This suggests that people, regardless of their age, need a certain amount of time to adapt to an environmental change to appreciate its consequences in terms of effort and confidence. Overall, these results suggest that the ability to track effort and confidence online is preserved even in older age, at least for individuals with typical hearing. Note that older participants easily understood the task, corroborating the idea that metacognitive tracking skills are preserved in this population.

As anticipated in the previous paragraph Giovanelli and colleagues (2023) asked participants to rate their confidence after each experiential segment. They observed a difference between young and older adults, revealing that the latter reported lower confidence values. This result is in contrast with our observation in this study as we did not find any effect of the group considering confidence, not even directly measuring the average of the confidence values reported during the story. The incongruency between these two results could be partially attributed to the response type (online vs. after experience) but also to the focus of the two tasks. In the case of Giovanelli et al., they were asked to rate their confidence about the correctness of each word they reported (i.e., they were asked explicitly to rate their confidence with respect to achieve a specific request), whereas in our study, we did not explicitly link confidence to a precise request but asked them to constantly judge their confidence compared to what they perceived across the story flow.

Nevertheless, this difference in confidence rating did not indicate a different sensitivity in modifying one's confidence depending on the listening condition between young and older adults.

An additional discussion point is how we can be certain that the measured responses truly reflect the participants' ability to perceive their internal states and are not simply a consequence of noticing SNR variations. This is a common issue in many tasks that collect both objective and subjective measures. In our study, this issue can be partially addressed by considering that participants were instructed to focus on their internal states and explicitly pay attention to this aspect rather than to noise fluctuations. They were also told not to change their ratings if they did not perceive any fluctuations in effort or confidence, and we did not emphasize the presence of varying noise levels in the instructions, only mentioning that there might be some background noise. Furthermore, we aimed to keep the signal easily distinguishable from the noise during the task (intelligibility was calculated to be always above 90 %). In this way, the SNR variations had minimal impact on comprehension and could be less prominently used as the main information for judgment changes. Plus, while effort ratings fluctuated in the same direction as the variation in the volume of the noise, confidence changed in the completely opposite direction. This suggests that it is more parsimonious to interpret these data as participants focusing on their internal state rather than on variations in the physical stimulus. Additionally, we found that the maximum correlation value (R Maximum) occurred with a delay of about 4 s (Time R Maximum) relative to the SNR variation. This suggests that participants were more focused on the fluctuations of their internal states. If they had been focusing on the SNR, we would have expected a shorter delay. We believe that future studies could experimentally test this hypothesis by explicitly shifting the focus of attention to detecting changes in SNR to see if different results are obtained. If a shorter delay is observed, it could reflect that it is easier to detect SNR variations than internal sensations.

As far as concern the effort rating, again, we did not observe any difference by considering the indices analysed, with the only exception of the mean value of perceived effort. On average, the values reported by the older adults (2.03 ± 0.59) were lower than those of the young (2.38 ± 0.60). This difference was unexpected and we noted that it did not survived when we corrected for multiple comparison. We speculate that older adults may have rated it less tiring and slightly easier than young adults, likely relying on their accumulated knowledge and direct experiences in challenging situations (Fleming, 2024). Upon visually inspecting Fig. 3, we noticed that older adults tended to report lower effort levels, particularly at the beginning of the story, compared to younger adults. This observation suggests an alternative hypothesis: older adults may have intentionally kept their initial ratings low as a strategy to assess their capacity for exertion, allowing them to calibrate their level of effort more effectively as the story progressed. However, this result appears to be in contrast with some of the literature on effort. Studies that investigated differences in effort between young and older adults in listening-in-noise tasks observed that older reported more effort than young participants. These studies have used physiological indices such as pupil size (Zekveld et al., 2011) or dual-task paradigms (Degeest et al., 2015). Interestingly, as highlighted in the study by Zekveld and collaborators, the cost of age emerges especially in the most difficult conditions. Given that the task proposed in our study is not particularly challenging throughout the entire course of the evaluation (the story is almost always intelligible > 90 %), older adults probably did not experience a sufficiently demanding listening environment to observe such age differences. Furthermore, a recent study by McGarrigle and Mattys (2023) investigating the interaction between sensory-processing sensitivity, listening fatigue, and effort revealed that older adults found a dichotic listening task more effortful but reported lower overall fatigue. This evidence (see also Herrmann & Johnsrude, 2020; Pichora-Fuller et al., 2016) suggests a need to further explore the relationship between effort and fatigue and to consider other factors such as engagement and mood experienced during listening situations. Understanding these mechanisms could help clarify the awareness of fluctuations in feelings such as effort and confidence. This could inspire future studies and contribute to describing the differences in hearing in noise between younger and older individuals.

Another possibility of this effort difference between older and young adults might be found in the nature of our task. Previous studies analysing age differences in listening found that older adults might compensate for auditory losses by an increased reliance on semantic context (Rogers, Jacoby & Sommers, 2012). In other words, older adults are more inclined to perceive words suggested by context, even at the cost of committing false hearing (hearing one word for another and being absolutely sure to be right). As our stimuli are stories with a meaning, it is possible that older adults keep a perception of hearing well and with less effort because they hear words covered by noise through a top-down semantic process, most likely maintaining a perception of continuous understanding. As our task did not require an exhaustive assessment of performances, we cannot exclude that older adults actually fail to really hear words or commit false hearing. In any case, the heavier reliance on context may account for part of the age differences observed in this task.

A further aspect of the investigation of this study concerns story comprehension. Again, we observed that both groups equally struggled to remember the contents of the story. This result is somewhat unexpected. In fact, if the older adults had used more resources than the young in carrying out the task of tracking their internal states, fewer resources should have been left to memorize the contents of the story, and therefore, we should have found a difference between the two groups. This result would instead seem to suggest a similar amount of cognitive resources used to carry out the test by both populations. This corroborates the idea of a preserved metacognitive capacity with the passage of age. However, this interpretation is only a speculation and requires further investigation. Future studies could adopt different methodological strategies to test resources distribution (e.g., dual-task paradigm).

5. Conclusion

To conclude, in this study, we introduced a novel approach to metacognitive tracking, which assesses individuals' continuous, online evaluations of their internal states such as effort and confidence. This subjective paradigm adds depth to existing listening effort tests, emphasizing the need to consider first-person accounts alongside traditional objective measures. The challenge of listening in noise engages a wide range of functions and resources, activating the person as a whole and leading to the perception of uncertainty,

effort, and discomfort. This translates into alterations in behavior, is reflected in physiological parameters such as pupil dilation, and is clearly observed in brain activity, involving various areas and circuits beyond the auditory pathways (Peelle, 2018). Our behavioral approach complements existing techniques and offers insights into the listening experience. This has potential implications for clinical interventions, such as hearing-aid tuning, and understanding the impact of noise on quality of life.

These results suggest the potential to apply this methodological approach to other perceptual domains or the field of metacognition related to memory. They also expand the definition of metacognition by introducing the concept of metacognitive tracking ability (see Fleming, 2024). We showed that both young and older adults can consistently monitor their internal states in response to environmental changes, such as noise fluctuations. Although it is well-documented that older adults generally encounter more difficulties than younger individuals in noisy environments—evidenced by a general broader brain activity, involving also attention and working memory circuits (Kuchinsky & Vaden, 2020)—our findings show that older adults are still capable of monitoring their internal states under these challenging conditions. This ability could be a strength, suggesting an intact neural system and offering a resource for developing strategies to help them manage everyday perceptual difficulties (Hertzog & Dunlosky, 2011; Dehaene, 2020; Stine-Morrow et al., 2008). Future research could explore this further to understand the dynamic balance between internal state monitoring and speech recognition in older adults, and to find ways to strengthen the link between self-awareness and performance, especially in challenging listening situations.

CRediT authorship contribution statement

Chiara Valzolgher: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alessandra Federici:** Writing – review & editing, Writing – original draft, Formal analysis. **Elena Giovanelli:** Writing – review & editing, Writing – original draft, Resources. **Elena Gessa:** Writing – review & editing, Writing – original draft, Resources. **Davide Bottari:** Writing – review & editing, Supervision, Formal analysis. **Francesco Pavani:** Writing – review & editing, Supervision, Resources.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 to enhance the readability of certain sentences written by non-native English speakers. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets generated and analyzed during the current study are available in OSF repository, and retrieved from osf.io/86jku.

Acknowledgments

We are grateful to all the participants in the study and we thank Marta Fantoni for consultancy in statistical analysis.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2024.103747.

References

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences, 21*(8), 607–617. https://doi.org/10.1016/j.tics.2017.05.004

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory- cognitive system supports speech-in-noise perception in older adults. *Hearing Research, 300*, 18–32. https://doi.org/10.1016/j.heares.2013.03.006

Barajas, M. C. O., Guevara, R., & Gervain, J. (2021). The origins and development of speech envelope tracking during the first months of life. *Developmental Cognitive Neuroscience, 48*, Article 100915. https://doi.org/10.1016/j.dcn.2021.100915

Bednaya, E., Mirkovic, B., Berto, M., Ricciardi, E., Martinelli, A., Federici, A., … Bottari, D. (2022). Early visual cortex tracks speech envelope in the absence of visual input. *BioRxiv*, 2022-06.

Birren, J. E., & Fisher, L. M. (1992). Aging and slowing of behavior: Consequences for cognition and survival. In *Psychology and Aging: Nebraska Symposium on Motivation* (Vol. 39, pp. 1–37). Lincoln, NE: University of Nebraska Press.

Bowl, M. R., & Dawson, S. J. (2019). Age-related hearing loss. *Cold Spring Harbor Perspectives Medicine, 9*(8), Article a033217. https://doi.org/10.1101/cshperspect.a033217

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America, 111*(6), 2801–2810. https://doi.org/10.1121/1.1479152

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics, 77*(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

Degeest, S., Keppler, H., & Corthals, P. (2015). The effect of age on listening effort. *Journal of Speech, Language, and Hearing Research, 58*(5), 1592–1600. https://doi.org/10.1044/2015_JSLHR-H-14-0288

Dehaene, S. (2020). *How We Learn: Why Brains Learn Better Than Any Machine for Now*. Viking.

Di Liberto, G. M., O'sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology, 25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(4), 277–287. https://doi.org/10.1027/1016-9040.13.4.277

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive –developmental inquiry. *The American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology, 75*(1), 241–268. https://doi.org/10.1146/annurev-psych-022423-032425.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Fostick, L., & Schneider, B. A. (2022). Editorial: Age-related changes in auditory perception. *Frontiers in Psychology, 13*, Article 986586. https://doi.org/10.3389/fpsyg.2022.986586

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology, 43*(2), 85–99. https://doi.org/10.1080/14992020400050014

Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., & Pavani, F. (2021). Unmasking the difficulty of listening to talkers with masks: Lessons from the COVID-19 pandemic. *i-Perception, 12*(2), Article 2041669521998393. https://doi.org/10.1177/20416695219983

Giovanelli, E., Valzolgher, C., Gessa, E., Rosi, T., Visentin, C., Prodi, N., & Pavani, F. (2023). Metacognition for hearing in noise: A comparison between younger and older adults. *Aging Neuropsychology and Cognition, 1–22*. https://doi.org/10.1080/13825585.2023.2281691

Gagne, J. P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in hearing, 21*, Article 2331216516687287.

Gosselin, P. A., & Gagne, J. P. (2011). Older adults expend more listening effort than young adults recognizing speech in noise. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2010/10-0069)

Glyde, H., Hickson, L., Cameron, S., & Dillon, H. (2011). Problems hearing in noise in older adults: A review of spatial processing disorder. *Trends in Amplification, 15*(3), 116–126. https://doi.org/10.1177/1084713811424885

Herrmann, B., & Johnsrude, I. S. (2020). A model of listening engagement (MoLE). *Hearing Research, 397*, Article 108016. https://doi.org/10.1016/j.heares.2020.108016

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20*(3), 167–173. https://doi.org/10.1177/0963721411409026

Hertzog, C., Dunlosky, J., Powell-Moman, A., & Kidder, D. P. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*(2), 209–225. https://doi.org/10.1037/0882-7974.17.2.209

Howard, C. S., Munro, K. J., & Plack, C. J. (2010). Listening effort at signal-to-noise ratios that are typical of the school classroom. *International Journal of Audiology, 49*(12), 928–932. https://doi.org/10.3109/14992027

Kaakinen, J. K., & Simola, J. (2020). Fluctuation in pupil size and spontaneous blinks reflect story transportation. *Journal of Eye Movement Research, 13*(3). https://doi.org/10.16910/jemr.13.3.6

Klimovich-Gray, A., Di Liberto, G., Amoruso, L., Barrena, A., Agirre, E., & Molinaro, N. (2023). Increased top-down semantic processing in natural speech linked to better reading in dyslexia. *NeuroImage, 273*, 120072.

Koelewijn, T., de Kluiver, H., Shinn-Cunningham, B. G., Zekveld, A. A., & Kramer, S. E. (2015). The pupil response reveals increased listening effort when it is difficult to focus attention. *Hearing Research, 323*, 81–90.

Kuchinsky, S. E., & Vaden, K. I. (2020). Aging, hearing loss, and listening effort: Imaging studies of the aging listener. *Aging and Hearing: Causes and Consequences*, 231–256.

Laeng, B., & Alnaes, D. (2019). Pupillometry. *Eye Movement Research: An Introduction to its Scientific Foundations and Applications*, 449–502.

Lakatos, P., Gross, J., & Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. *Current Biology, 29*(18), R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

Lemke, U., & Besser, J. (2016). Cognitive load and listening effort: Concepts and age-related considerations. *Ear and Hearing, 37*, 77S–84S. https://doi.org/10.1097/AUD.0000000000000304

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

McGarrigle, R., Knight, S., Rakusen, L., Geller, J., & Mattys, S. (2021). Older adults show a more sustained pattern of effortful listening than young adults. *Psychology and Aging, 36*(4), 504. https://doi.org/10.1037/pag0000587

McGarrigle, R., & Mattys, S. (2023). Sensory-processing sensitivity predicts fatigue from listening, but not perceived effort, in young and older adults. *Journal of Speech, Language, and Hearing Research, 66*(2), 444–460. https://doi.org/10.1044/2022_JSLHR-22-00374

Mick, P., Kawachi, I., & Lin, F. R. (2014). The association between hearing loss and social isolation in older adults. *Otolaryngology and Head and Neck Surgery, 150*(3), 378–384. https://doi.org/10.1177/01945998135180

Nadol, J. B., Jr (1993). Hearing loss. *The New England Journal of Medicine, 329*(15), 1092–1102. https://doi.org/10.1056/NEJM199310073291507

Nunnally, J. C., Knott, P. D., Duchnowski, A., & Parker, R. (1967). Pupillary response as a general measure of activation. *Perception & Psychophysics, 2*, 149–155. https://doi.org/10.3758/BF03210310

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences, 23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., & Kramer, S. E. (2017). Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review. *Ear and Hearing, 38*(3), 267. https://doi.org/10.1097/AUD.0000000000000396

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition, 28*, 151–160. https://doi.org/10.1016/j.concog.2014.06.007

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing, 39*(2), 204–214. https://doi.org/10.1097/AUD.0000000000000494

Pichora-Fuller, M. K. (1997). Language comprehension in older listeners. *Journal of Speech-Language Pathology and Audiology, 21*(2), 125–142.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., 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 and Hearing, 37*(1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Puglisi, G. E., Warzybok, A., Hochmuth, S., Visentin, C., Astolfi, A., Prodi, N., & Kollmeier, B. (2015). An Italian matrix sentence test for the evaluation of speech intelligibility in noise. *International Journal of Audiology, 54*(Suppl 2), 44–50. https://doi.org/10.3109/14992027.2015.1061709

Rabini, G., Lucin, G., & Pavani, F. (2020). Certain, but incorrect: On the relation between subjective certainty and accuracy in sound localisation. *Experimental Brain Research, 238*, 727–739. https://doi.org/10.1007/s00221-020-05748-4

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging, 27*(1), 33. https://doi.org/10.1037/a0026231

Saksida, A., Ghiselli, S., Picinali, L., Pintonello, S., Battelino, S., & Orzan, E. (2022). Attention to speech and music in young children with bilateral Cochlear implants: A Pupillometry study. *Journal of Clinical Medicine, 11*(6), 1745. https://doi.org/10.3390/jcm11061745

Santangelo, G., Siciliano, M., Pedone, R., Vitale, C., Falco, F., Bisogno, R., Siano, P., Barone, P., Grossi, D., Santangelo, F., & Trojano, L. (2015). Normative data for the Montreal Cognitive Assessment in an Italian population sample. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, 36*(4), 585–591. https://doi.org/10.1007/s10072-014-1995-y

Sim, J., Brown, F. L., O'Connell, R. G., & Hester, R. (2020). Impaired error awareness in healthy older adults: an age group comparison study. *Neurobiology of Aging, 96*, 58–67. https://doi.org/10.1016/j.neurobiolaging.2020.08.001

Stine-Morrow, E. A., Soederberg Miller, L. M., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging, 23*(1), 131–153. https://doi.org/10.1037/0882-7974. 23.1.131

Sussman, E. S. (2017). Auditory scene analysis: An attention perspective. *Journal of Speech, Language, and hearing research: JSLHR, 60*(10), 2989–3000. https://doi.org/10.1044/2017_JSLHR-H-17-0041

Tullis, J. G., & Benjamin, A. S. (2012). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging, 27*(3), 683–690. https://doi.org/10.1037/a0025838

Visentin, C., Pellegatti, M., Garraffa, M., Di Domenico, A., & Prodi, N. (2023). Individual characteristics moderate listening effort in noisy classrooms. *Scientific Reports, 13*(1), 14285. https://doi.org/10.1038/s41598-023-40660-1

Wendt, D., Hietkamp, R. K., & Lunner, T. (2017). Impact of noise and noise reduction on processing effort: A pupillometry study. *Ear and Hearing, 38*(6), 690–700. https://doi.org/10.1097/AUD.0000000000000454

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started, 2331216518800869 *Trends in hearing, 22.* https://doi.org/10.1177/2331216518800869.

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing, 32*(4), 498–510. https://doi.org/10.1097/AUD.0b013e31820512bb