

Supplemental Material S1. Experimental materials.

The participant was provided with URL links to the self-report measures, as well as individual links to each of the behavioral tasks. Participants completed the measures and tasks in the following order, unless a technical issue prevented it: 1) self-report measures, 2) forward digit span, 3) backward digit span, 4) choice reaction time, 5) Stroop color-word, 6) speech-in-noise. In the case of a technical issue, participants were asked to move onto the next task while the researcher resolved the potential issue. The participant was informed that they could take breaks between tasks but not during them and were asked to complete all questionnaires and tasks on the same day where possible.

Social Isolation

A composite social isolation measure was computed by z-standardizing the total scores within each questionnaire, and then calculating the mean.

- *Lubben Social Network Scale Version 6 (LSNS-6)*

The LSNS-6 is a six-item questionnaire used to assess an individual's perception of social support available to them and frequency of contact with their social networks. An example question is "how many relatives did you see or hear from at least once a month?"

Participants responded using a 6-point Likert scale containing the following choices: none, one, two, three or four, five to eight, or nine or more. This questionnaire is deemed to have high reliability and validity in older adult populations (Lubben et al., 2006).

- *UCLA Loneliness Scale Version 3 (UCLA-LS3)*

The UCLA-LS3 is a 20-item questionnaire used to assess feelings of loneliness and disconnect from others. An example question is "how often do you feel alone?" and "how often did you feel that you lacked companionship?" Participants respond using a 4-point Likert scale containing the following choices: never, rarely, sometimes, or always. The questionnaire has been shown to have high reliability and validity, across age ranges (Russell, 1996).

Depression

- *Beck Depression Inventory (BDI-I)*

The BDI-I is a 21-item questionnaire used to evaluate the severity of depressive symptoms experienced by a participant over the previous week. For each item, the participant selected one of four statements which range in intensity, with each of the statements scored on a 0–3 scale. For example, I do not feel sad (0), I feel sad (1), I am sad all the time and I can't snap out of it (2), or I am so sad or unhappy that I can't stand it (3). The questionnaire has high reliability and validity (Beck et al., 1988; Richter et al., 1998).

Auditory and Lifestyle Engagement

A novel 10-item questionnaire was designed for this study that assessed subscales of auditory and lifestyle engagement. The overall reliability of the engagement questionnaire across all assessed factors (10-items) was acceptable ($\alpha = .63$) (Cronbach, 1951; Nunnally & Bernstein, 1994). The individual alpha for each of the two sub-factors was, $\alpha = .58$ for total auditory engagement, and $\alpha = .43$ for lifestyle engagement.

Items for the auditory and lifestyle engagement questionnaire

Introductions for participants:

For this questionnaire, please think about an average week in the current month. If it is the beginning of the month and a week has not yet passed, then please think of an average week in the previous month. There are 10 questions regarding certain activities, and for

each you will need to indicate for how many hours you engaged with that activity during the average week. Participants used a slider from 0-50 to indicate their answer.

Items 1-7 assessed auditory engagement

1. On average this month, how much time per week did you spend doing any in-person social activities with other people outside your home? Please include activities with any persons you live with. Please include socializing at work, and any activity that involved in-person spoken communication.
2. On average this month, how much time per week did you spend doing any in-person social activities with other people outside your home? NOT including activities with any persons you live with. Please include socializing at work, and any activity that involved in-person spoken communication.
3. On average this month, for how many hours per week did you participate in-person in social hobby groups (knitting, book clubs etc.)?
4. On average this month, for how many hours per week did you participate in hobby activity groups (knitting, book clubs etc.) via. an online platform?
5. On average this month, for how many hours per week did you speak with other people over the phone, or online (i.e., phone, or video calls with friends and family)?
6. On average this month, for how many hours per week did you engage in a listening activity that involved watching TV, films, or videos?
7. On average this month, for how many hours per week did you engage in a listening activity that involved listening to the radio, audio books, podcasts, or listening to (or playing) music (including singing)?

Items 8-10 assessed lifestyle engagement

8. On average this month, for how many hours per week did you engage in a solitary hobby (such as, gardening, crafts)?
9. On average this month, for how many hours per week did you engage in in any sport or exercise outside your home with other people?
10. On average this month, for how many hours per week did you engage in in any sport or exercise outside your home alone?

- ***Auditory Engagement Subscale***

Seven items which assessed auditory engagement measured how much time participants spent doing auditory activities across active (including in-person and online communicative activities) and passive listening domains (including non-communicative activities like listening to audiobooks). The auditory questionnaire items were weighted based on the level of auditory engagement they were designed to assess. The total score obtained from the summed responses to items 1–3 for in-person communication, was multiplied by 0.3. The total score from items 4–5 for online communication, was multiplied by 0.2, and for items 6–7 for non-communication activities, it was multiplied by 0.1. The decision to employ these weighting was made a-priori and preregistered and was designed to ensure that activities which involved greater in-person communication were given greater importance. The resulting scores from the 7 items were totaled to provide an auditory engagement score, wherein higher scores indicate greater auditory engagement.

- ***Lifestyle Engagement Subscale***

Three items measured the time participants estimated that they spent engaged in various lifestyle activities such as hobbies, or sports. The total score obtained from the summed responses to the three items provided a total lifestyle engagement score. Higher scores indicate greater lifestyle engagement.

Cognitive Function

A composite score of global cognition was obtained following z-standardization and then averaging the score on the following cognitive assessment tasks. Composite scores may be advantageous for consistent reliability (Waters & Caplan, 2003).

- ***Forward Digit Span***

The Forward Digit Span (e.g., Wechsler, 1997) was used to assess short-term memory. Participants were presented with eight sets of number sequences containing two sequences per set, in order of difficulty. The sequence length ranged from two digits in set one to nine digits in set eight. In a trial, participants saw a fixation cross (displayed for 1 sec), followed by each number in the sequence (1 sec for each number), and then a response screen, where they were asked to type the number sequence. After the response, participants saw a blank screen for 1 sec before the next trial began. The task ended if two sequences in a set were recalled incorrectly. The number of correctly recalled sequences was totaled (range 0–16), with higher scores indicating better short-term memory performance.

- ***Backward Digit Span***

The Backward Digit Span (e.g., Wechsler, 1997) was used to assess working memory. Participants were presented with seven sets of number sequences containing two sequences per set, in order of difficulty. The sequence length ranged from two digits in set one to eight digits in set seven. In a trial, participants were presented with a fixation cross (for 1 sec), followed by each number in the sequence (1 sec for each number), and then a response screen, where they were asked to type the number sequence in the reverse order. After the response, participants saw a blank screen for 1 sec before the next trial began. The task ended if two sequences in a set were recalled incorrectly. The number of correctly recalled sequences was totaled (range 0-14), with higher scores indicating better working memory performance.

- ***Deary-Liewald Choice Reaction Time***

The Deary-Liewald Choice Reaction Time (Deary et al., 2011) was used to assess processing speed. Participants were presented with four on-screen squares in a horizontal line in a randomized order. In a trial, a target 'x' appeared in one of the four squares, and the participant used their number keys to indicate which box the target appeared in, where 1 indicated the box furthest left, and 4 indicated the box furthest right. The inter-trial interval varied between 1 and 3 secs, and there were 40 trials in total. The time at which the target position was identified was used to calculate a mean reaction time. In order to be consistent with the other cognitive measures, the mean was reversed prior to calculating the global cognition composite so that better reaction time performance was indicated by higher.

- ***Stroop Color-Word***

This Stroop Color-Word task (Stroop, 1935) task was used to assess executive function. The task consisted of three conditions each containing 48 trials (words only, colors only, or color-words), resulting in a total of 144 trials which were presented in condition blocks. In the words-only condition, participants were presented with a fixation cross (1 sec) followed by a word (either RED, GREEN, YELLOW, or BLUE) in white text on a grey background. The participant was instructed to recall the word they saw by pressing either the 'R', 'G', 'Y', or 'B' key, respectively. In the colors-only condition, participants were presented with the repeated letter X in the colors of red, green, yellow, or blue. Participants were instructed to recall the color of the Xs by pressing either the 'R', 'G', 'Y', or 'B' key. In the color-words condition, participants were presented with the color word (either RED, GREEN, YELLOW, or BLUE) printed in incongruent colored text (e.g., the word BLUE printed in red color). Participants were instructed to recall the color of the text, not the word itself, by pressing either the 'R', 'G', 'Y', or 'B' key. An interference score was

calculated using a method adapted from Golden (1978) using the following formula, where higher scores indicate better ability to inhibit interference:

$$\text{Predicted color-words score} = 48 \div (((48 \times \text{words only total}) + (48 \times \text{colors only total})) \div (\text{words only total} \times \text{colors only total}))$$

Behavioral Auditory Function

- Speech-in-Noise

The Speech-in-Noise test was used to assess objective speech perception ability. The test scoring method was based on the formula employed in the BKB-SIN (Etymotic Research). This scoring formula is derived from the Tillman-Olsen method (Tillman & Olsen, 1973), and was adapted for this online task to estimate the SNR required for a person to identify 50% of target words correctly (SNR-50). This calculation is based on that used for calculating spondee thresholds in a speech-in-noise task in which the SNR increases in 2 dB steps and two key words need to be identified per trial (BKB-SIN Manual, Etymotic Research). The calculation was adapted to account for the five key words per 3 dB step in this task, wherein: 21 refers to the starting SNR level; 1.5 is half the step size; 2 is the number of additional key words per trial above the step size; Y is the number of SNR blocks where the participant scored higher than 2; and A is the total number of words correctly identified across all trials. A higher SNR50 indicates poorer performance:

$$\text{SNR-50} = (21 + 1.5 + (2 \times Y)) - A$$

Subjective Auditory Function

- Speech and Spatial Qualities of Hearing Scale (SSQ-12)

This 12-item questionnaire assessed subjective hearing ability (Noble et al., 2013). Participants responded on a 10-point Likert scale, where 0 indicated *very poor hearing ability* and 10 indicated *perfect hearing ability*. The scores were averaged over all items and then reversed poorer hearing ability was indicated by higher scores, in line with the objective measure.

References

- Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8(1), 77–100. [https://doi.org/10.1016/0272-7358\(88\)90050-5](https://doi.org/10.1016/0272-7358(88)90050-5)
- Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavior Research Methods*, 43(1), 258–268. <https://doi.org/10.3758/s13428-010-0024-1>
- Golden, C. J. (1978). Stroop Color and Word Test: A manual for clinical and experimental uses. *Chicago: Stoelting*.
- Lubben, J., Blozik, E., Gillmann, G., Iliffe, S., von Kruse, W. R., Beck, J. C., & Stuck, A. E. (2006). Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. *The Gerontologist*, 46(4), 503–513. <https://doi.org/10.1093/GERONT/46.4.503>
- Noble, W., Jensen, N. S., Naylor, G., Bhullar, N., & Akeroyd, M. A. (2013). A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: the

SSQ12. *International Journal of Audiology*, 52(6), 409-412.

<https://doi.org/10.3109/14992027.2013.781278>

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill, New York

Richter, P., Werner, J., Heerlein, A., Kraus, A., & Sauer, H. (1998). On the validity of the Beck Depression Inventory. A review. *Psychopathology*, 31(3), 160–168.

<https://doi.org/10.1159/000066239>

Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. *Journal of Personality Assessment*, 66(1), 20–40.

https://doi.org/10.1207/S15327752JPA6601_2

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>

Tillman, T. W., & Olsen, W. O. (1973). Speech audiometry. In J. Jerger (Ed.), *Modern developments in audiology* (pp. 37–74). New York: Academic Press.

Waters, G. S., & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, & Computers*, 35(4), 550–564.

<https://doi.org/10.3758/BF03195534>

Wechsler, D. (1997). Wechsler Adult Intelligence Scale (WAIS-3R). *The Psychological Corporation*. San Antonio, Texas.

Supplemental Material S2. Details of excluded participants and missing data replaced by mean imputation.

If any potential participants did not meet the inclusion criteria as stated in the Participants section, or if potential participants reported that they did not have access to the advised technical equipment (laptop or PC and earphones or headphones) they were unable to participate. After completing an initial screening survey: 4 were excluded due to being multilingual; 2 due to not being right-handed; 1 due to scoring too highly on the IQ-CODE cognitive decline screener; 2 due to having a diagnosis of a speech or language disorder; and 2 due to having a diagnosis of a neurological or psychiatric disorder. Further, if participants reported severe tinnitus or having severe hearing loss, to the point that they could not hear without the assistance of hearing aids, they were advised that they were unable to participate as part of the study required an unaided hearing task.

In the younger adult group, data were missing and thus replaced through mean imputation for the backward digit span in one participant, and the speech-in-noise task across four participants. In the older adult group, data were missing and thus replaced through mean imputation for all self-report measures across two participants, the backward digit span in one participant, the Deary-Liewald choice reaction time task across two participants, the Stroop color-word task across one participant, and the speech-in-noise task across three participants. In total, means were inputted across various measures for 4 participants from the younger adult group, and 8 participants in the older adult group.