Title: fNIRS, Listening Effort, and Motivation

NCT number: NCT05893992

Document date: May 5th 2025

Design

The experiment was a one-factor (two-level) repeated measures within-subjects design. The independent variable was hearing aid program, comprising (1) standard and (2) DNN-based programs as factor levels. The dependent variables, like our previous design [17], were listening accuracy, subjective listening effort, and HbDiff. Listening accuracy was measured as percent of correctly repeated words. Listening effort was measured on a 7-point Likert scale [27] (1-no effort, 2-very little effort, 3-little effort, 4-moderate effort, 5-considerable effort, 6-much effort, 7-extreme effort). HbDiff was measured using a prefrontal fNIRS system consisting of 6 active- and 2 short-channel optodes positioned across the prefrontal cortex.

Test materials

Stimuli consisted of sequences of up to five R-SPIN sentences presented in noise. Sequences, rather than individual sentences, were used to leverage the superpositioning of the response functions for each sentence in the sequence to enhance the overall temporal summation, thus increasing the SNR of the overall hemodynamic response [28–31]. Sequences varied in number of sentences so that participants would be unable to anticipate which word to repeat, forcing them to attend to the entire sequence. Sequences of three or more sentences were comprised of low-context sentences and were 18 or more seconds in duration. The shorter "decoy" sequences were comprised of one or two high-context sentences and were only included to create the impression that some sequences would contain only that number of sentences. The shorter high-context sequences' data were not used for analysis. Noise consisted of multitalker babble and was gated on and off with each sentence. Speech was presented at individualized SNRs relative to the noise level, which was maintained at 70 dBA.

Equipment

Stimulus playback

The experiment was conducted in a double-walled, sound-attenuated booth. The participant was seated in the center of the booth. The listening paradigm was administered using a custom MATLAB script (version R2021b) running on a Windows computer located outside the booth, which controlled stimulus playback, recorded participant responses, and transmitted stimulus onset and offset markers to the fNIRS recording software. The Windows computer was connected to an RME Fireface UCX sound card, which routed the stimuli to two self-powered Genelec 8030C speakers positioned at 0 and 180 degrees relative to the participant's chair, each placed 1.3 meters away from the participant's head. speech was presented from the front speaker while noise was presented from the rear speaker.

fNIRS device

Cerebral oxygenation was measured using an Octamon continuous-wave fNIRS device (Artinis Medical Systems, Netherlands) with six active optodes and two-short channel optodes, consisting of eight LED diode emitters (wavelengths: 760 and 850 nm) and two photodiode detectors, sampled at 10 Hz. The light emitters and detectors were spaced approximately 3.5 cm apart, allowing penetration of about 1.75 cm into the cortex. The device was positioned over the participant's prefrontal cortex. The base of the fNIRS headband was aligned just above the nasion, ensuring the lower medial light emitters were symmetrically positioned around the FpZ coordinate [32]. Recordings were transmitted via Bluetooth to the OxySoft software (version 3.2.70) on a Windows computer. The data were initially saved in .oxy4 format and later converted to .snirf format for analysis using MATLAB-integrated Homer3 scripts.

Hearing aids

The hearing aids used were Phonak Audéo I90-Sphere receiver-in-the-canal devices and coupled to participants' ears using occluding silicone power domes. The hearing aids were programmed using Phonak's proprietary gain formula and fine-tuned individually to NAL-NL2 targets by a licensed audiologist. Phonak's "Calm situation" and "Spheric speech in loud noise" programs, with all program features at default values, were used for the standard- and DNN-listening conditions, respectively. The standard-listening program is the hearing aid's default start-up program and implements omnidirectional microphones, speech enhancement and fast-acting compression; features common across manufacturers' standard-listening programs. The DNN-listening condition implements a combination of directional microphones DNN-based noise management, and slow-acting compression. Hearing aids were worn for the full duration of testing.

Procedure

SNR-50 search

Participants first determined the SNR at which they could correctly repeat 50% of the presented words (SNR-50). Individual SNR-50 measurements calibrated the "difficulty" of the task such that participants would perform equally despite differences in thresholds or other characteristics. Participants wore the hearing aids in the standard-listening program during the SNR-50 search procedure. This procedure followed modified hearing-in-noise-test (HINT) SNR adaptation rules and the manual adaptive level controls from the Oldenburg Matrix Test. In each SNR-50 run, participants listened to 20 low-context R-SPIN sentences and repeated the final word of each. The SNR was adjusted dynamically - decreasing following a correct response and increasing after an incorrect one. The noise level was maintained at 70 dBA and the speech level varied systematically. The starting SNR was 8 dB. For the first five sentences, the SNR was adjusted in ±4 dB steps, while the remaining sentences were adjusted in ±2 dB increments. The SNR-50 was calculated as the mean SNR from sentences 12 to 20, plus the estimated SNR after the final trial. Each run consisted of 20 sentences randomly selected from the 25 available per R-SPIN list, with sentences from only one list used per run. Participants completed a total of three runs, with the average SNR-50+2 dB of the last two runs used for testing. The additional 2 dB were added to increase performance above chance so that participants would answer correctly more frequently and feel more motivated during testing. Throughout the task, the experimenter recorded participants' responses in real time to assess accuracy and trigger the next sentence. Across participants, the average SNR-50 was -0.6 dB (SD = 3.9).

fNIRS recording

The experimenter then placed the fNIRS device on the participant and paired it with the Oxysoft recording software. The signal quality was verified using the software's signal quality index algorithm, ensuring that the fNIRS recording SNR was above the noise floor and not saturated by ambient lighting. If the signal quality was insufficient, the experimenter adjusted the device fit.

Next, the experimenter provided the instructions for the main study task to the participants. The experimenter advised participants that they would hear sequences of sentences. Participants were instructed to repeat the last word of the last sentence in the sequence and then provide the number corresponding to their average listening effort rating for all the sentences in the sequence. Participants were then told that there would be a 30-second rest period between sequences and were asked to minimize movement and mind-wandering during baseline periods and testing. The experimenter was able to postpone the start of the following sequence (therefore elongating the baseline recording) in case the participant moved during the rest period.

The study task consisted of testing blocks, with each block containing a total of eight events or sentence sequences: six sequences of three or more low-context sentences and two "decoy" sequences of one or two high-context sentences. Each sequence was separated by a 30-second silent baseline recording period. Each block comprised sentences taken from a single R-SPIN list not used in the SNR-50 task. Altogether, participants completed five blocks: a practice block and four test blocks. The practice block was completed using the standard-listening program and two of the remaining four test blocks were assigned to either the standard- or DNN-listening conditions, for a total 16 sequences per condition. Test block order and R-SPIN list/block assignment was randomized between participants.

Analysis

fNIRS data processing

The fNIRS processing pipeline resembled our previous study's pipeline, which was adapted from Zhou et al., and was entirely implemented using MATLAB. The exception between our studies was that the current design incorporated short-channel subtraction. The fNIRS processing occurred as follows:

1. **Removal of step-like noise.** Brief losses of contact between the fNIRS headband and skin might produce abrupt steps in the fNIRS morphology and can negatively impact signal quality. These steps were removed by calculating the derivative of each channel's intensity time series, and absolute values greater than two SDs over the derivative's mean were zeroed. Channels were recalculated using the cumulative sum of the updated derivatives.

2. **Exclude channels of poor quality.** Signal quality was assessed using the scalp coupling index (SCI). The SCI extracts the heartbeat from the two fNIRS wavelengths by bandpass filtering between 0.5 and 1.5 Hz and then correlating the two filtered wavelengths. A poor correlation suggests that the fNIRS measurements did not accurately capture the heartbeat and would therefore not contain other meaningful physiological responses. Channels with an SCI < 0.75 were rejected, resulting in 11.5% of channels being rejected across participants.
3. **Conversion of light intensity to optical density.** Refer to Hupper et al.
4. **Correction of motion artifacts.** Movements during testing risk physical displacement of the optodes from the participant's head, risking noise in the data. The wavelet decomposition method was used to correct motion artifacts for any wavelet coefficients outside an interquartile range of 0.1.
5. **Conversion of optical density to hemoglobin concentration using the Modified Beer—Lambert law.** Refer to Delpy et al.
6. **Band-pass filter (0.01 – 1.5 Hz).** Low-frequency noise (such as drift) and high-frequency noise (such as breathing or heart rate) was band-pass filtered from the fNIRS signal.
7. **Short-channel correction:** Each channel underwent short-channel correction such that a general linear model was fit to the channel's fNIRS signal with the timeseries of the nearest short-channel (based on Euclidean distance) as the regressor. The physiological noise measured by the short-channel was regressed out of each channel's output.
8. **Band-pass filter (0.01 – 0.09 Hz).** An additional band-pass filter was applied to the data to remove any physiological noise missed prior to short-channel subtraction.

9. **Averaging.** HbO and HbR were calculated for each sequence of sentences. The hemodynamic response was measured as the average of 10 seconds following each sequence onset to the end of each sequence minus a 5-second baseline average prior to the start of the sequence. Ten seconds was chosen between sequence onset and measurement onset to allow for the hemodynamic response to reach its peak. Hemodynamic responses were then averaged across all low-context sequences for a given condition.
10. **Data for analysis.** The fNIRS response used for statistical analysis was calculated as the cerebral oxygen exchange (HbDiff = HbO - HbR). Three subregions, each containing averages across two optodes, were used for analysis. Lateral regions contained two peripheral optodes for each respective side and the lower medial region contained the two medial optodes.

Statistical modelling

Aligned with previous literature, subjective listening effort ratings and listening accuracy scores were averaged across trials resulting in a mean listening effort value and a mean accuracy score per participant. Afterwards, mean listening accuracy scores were transformed to rationalized arcsin units to reduce clustering of scores near ceiling performance. To assess the effect of the hearing aid program on subjective listening effort and on listening accuracy, two paired samples t-tests were conducted. A standardized effect size was calculated for each test using Cohen's *d*.

Blood oxygenation data (HbDiff) were modelled using multi-level modeling (MLM) because of the clustered data structure (i.e., high levels of covariance) due to the repeated measurement design. MLM estimates random effects for each cluster to provide unbiased model estimates (standard errors, regression coefficients, p-values). Furthermore, MLM is robust against cases of missing data, such as rejected channels falling below the SCI threshold[46]. In contrast, an

Analysis of Variance (ANOVA) is a complete-case analysis where an entire participant's data would be removed if the participant had any missing data from rejected channels.

We conducted three MLMs with HbDiff as the outcome variable: an unconditional model and two a priori theory-driven models to assess the effect of the hearing aid program and whether the effect of the hearing aid program is moderated by subregions of the prefrontal cortex. The main predictor variable was the hearing aid program (standard vs. DNN) and an interaction term between the hearing aid program and prefrontal subregion (left lateral vs. lower medial vs. right lateral). Categorical variables (hearing aid program and prefrontal subregion) were dummy coded. Standard listening and the left lateral subregion were used as the reference group for the hearing aid program and prefrontal subregion, respectively. Simple slopes analysis was used to probe the interaction.

To assess which model best fit the data, a nested modeling approach was used. This approach adds predictor variables to the previous model and assesses if the additional predictors significantly improve the amount of variance explained by the model using a likelihood ratio test (LRT). The Akaike information criterion (AIC) was also used to assess model fit with lower AIC values indicating a better fit to the data.

All analyses were conducted in R version 4.4.2. MLMs were estimated using full information maximum likelihood (FIML) and conducted with the following packages: *lme4*, *lmerTest*[4], *performance* and *interactions*. Paired samples t-tests were conducted using *stats* and Cohen's *d* was calculated using *lsr*. All statistical analyses were two-tailed tests with alpha of .05.