Perceptual Training to Improve Listeners' Ability to Understand Speech Produced by Individuals With Dysarthria

NCT04897711

Protocol Summary and Statistical Analysis Plan

This file was created on 7/11/24
IRB clearance to complete this work was granted 3/26/21

Experimental Protocol

Participants attended a single research session, lasting about two hours, at the Motor Speech Disorders Lab (MSDLab) at Florida State University or the Human Interaction Lab at Utah State University. After providing informed consent, participants completed a brief demographic survey, followed by the perceptual training protocol. The participants then completed the Words-in-Noise (WIN) test and the cognitive assessments using the NIH Toolbox Sensation and Cognition Batteries. At the time of data collection, adherence to health and safety protocols for COVID-19 was observed, including social distancing, mask-wearing, temperature monitoring, and a health questionnaire relevant to COVID-19 exposure risks. Prior to initiating the perceptual training task, participants were fitted with headphones and presented with a short audio clip. They were asked to adjust the volume to a comfortable listening level, and this volume was maintained throughout the perceptual task.

Perceptual Training Protocol. The study utilized a lexically-guided perceptual training protocol composed of three phases: pretest, familiarization, and posttest. These phases were designed to evaluate initial intelligibility and intelligibility change post-training. Participants, equipped with headphones and seated in a listening carrel, were randomly assigned to one of the speaker-specific training conditions. Instructions for each phase of the task were displayed on the computer screen. For the pretest, participants were informed that the speech samples they would hear might be difficult to understand and were asked to transcribe each phrase (n = 20) after hearing it once. They were encouraged to make their best guess if uncertain. Following the pretest, the familiarization phase required participants to listen to the same speaker's audio recordings of the Grandfather Passage, segmented into 35 sentences or phrases, while following along the orthographic transcription of each segment presented on the screen. They were able to advance to subsequent segments at their own pace. The final posttest phase was identical in structure to the pretest, except participants were asked to transcribe 60 new phrases spoken by the same speaker. Each phase of the perceptual training protocol was self-paced and untimed. On average, participants took approximately 30 minutes to complete the perceptual training protocol.

Hearing and Cognitive Assessment with NIH Toolbox. The NIH Toolbox (Weintraub et al., 2013) was administered via an iPad Pro to assess hearing and cognitive abilities. The WIN test was administered from the Sensation Battery to estimate auditory signal processing under challenging listening conditions. The Cognitive Battery includes a variety of tasks that estimate specific cognitive-linguistic skills, including working memory, inhibitory control, cognitive flexibility, processing speed, and vocabulary knowledge. The specific tasks used to quantify each construct are detailed briefly in Table 3. The administration of these tasks adhered closely to the NIH Toolbox guidelines with slight modifications to maintain social distancing.

Table 1. Tasks used by the NIH Toolbox Cognition Battery to quantify the hearing and cognitive-linguistic predictors

Variable	**Description of the Task**
Hearing	The *WIN* test presents a series of monosyllabic words embedded in a background of steady-state speech babble noise. After each presentation, participants are instructed to focus on the target words and repeat them back as accurately as possible. The signal-to-noise ratio (SNR) starts at an easier level and progressively becomes more challenging, providing a gradient measure of speech recognition thresholds in noisy environments.
Working memory	To estimate working memory capacity, a *list sorting memory test* was used. In each trial, a series of pictures from two different categories (e.g., food and animals) is presented one at a time. At the end of the trial, participants are instructed to recall the names of the pictures from one

	category first, in size order (smallest to largest). Participants then complete the same task with the pictures from the other category.
Inhibitory control of attention	To obtain estimates of each participant's inhibitory control of attention, *a flanker test* was administered. In this task, a row of arrows is presented, and the participant is instructed to select the direction in which the middle arrow points (right or left). Distractor arrows sometimes point in the same or opposite direction.
Cognitive flexibility	A *dimensional change card sort test* was administered to measure cognitive flexibility. This task presents a pair of pictures that differ along two dimensions relative to a target picture. The participant is instructed to select the test picture that matches the target picture based on one dimension. After several trials, the participant sorts the images based on the other dimension.
Processing speed	*A pattern comparison test* yields estimates of processing speech. In this task, a pair of pictures is presented, and the participant is instructed to quickly determine whether the pictures are the same while trying not to make mistakes.
Vocabulary knowledge	A *picture vocabulary test* was administered to obtain estimates of vocabulary knowledge. Four pictures are presented, and the participant hears a word spoken aloud. The participant is instructed to select the picture that matches the spoken word.

Statistical Analysis Plan

Analyses for this project were stratified into two main analyses, based on the subsample: 1) younger listeners and 2) older listeners. Both plans are similar with transcript analyses with Autoscore and statistical analyses that include testing for changes from pretest ("initial intelligibility") and to posttest. In addition, we will perform predictive analyses to understand these patterns more fully.

Younger Listeners

Transcript Analysis.

Listener transcriptions from both the pretest and posttest phases will be evaluated for accuracy using Autoscore, an open-source computer-based tool for automated intelligibility scoring (http://autoscore.usu.edu; Borrie et al., 2019). The Autoscore settings were calibrated to score a word as accurately transcribed, not only when it matched the target word but also when variations were due to tense changes or pluralization. Additionally, homophonic errors and prevalent spelling mistakes were scored as accurate utilizing a built-in list of commonly misspelled words. Each listener's performance was quantified by calculating the percentage of words correctly transcribed (PWC) in both testing phases, yielding a pretest PWC score and a posttest PWC score for each listener. These PWC scores were used to quantify initial intelligibility (pretest PWC) and intelligibility improvement (posttest PWC after accounting for pretest PWC).

Statistical Analyses.

Initial analyses will investigate distributions of the variables, assess for baseline differences, test for changes from pretest to posttest, and calculate correlations between variables of interest. Descriptive statistics (means and standard deviations) will be calculated for all variables of interest (e.g., cognitive measures, intelligibility), and baseline differences between conditions were tested using Kruskal-Wallis Rank Sum tests. Paired-samples t-tests will be used to test for average intelligibility improvement from pretest to posttest with effect sizes estimated with Cohen's *d* for each speaker condition. Lastly, Pearson correlations will highlight relationships between each cognitive measure and their relationship with intelligibility (both initial and improvement) by speaker condition.

Predictive modeling—using elastic net regression models—will assess the ability of the cognitive measures to predict both initial intelligibility and intelligibility improvement. Model-specific hyper-parameters will be selected based on 10-fold cross-validation. Predictive models will be run separately for each speaker condition. For each speaker, two model specifications will be used: 1) main effects only and 2) main effects and all two-way interactions. The model predictions and variable importance metrics, as well as post hoc assessment of predictions, will inform on the nature of the important interactions. Predictive accuracy includes the root mean squared error (RMSE) and R^2 values for the cross-validated predictions. Variable importance will be extracted from each model using a permutation method wherein each variable is randomly permuted, and the difference in predictive accuracy between the observed variable and its randomly permuted counterpart was derived.

All analyses will be completed in the R statistical environment version 4.2.0 or higher.

Older Listeners

Transcript Analysis.

Listener transcriptions from both the pretest and posttest phases will be evaluated for accuracy using Autoscore, an open-source computer-based tool for automated intelligibility scoring (http://autoscore.usu.edu; Borrie et al., 2019). The Autoscore settings were calibrated to score a word as accurately transcribed, not only when it matched the target word but also when variations were due to tense changes or pluralization. Additionally, homophonic errors and prevalent spelling mistakes were scored as accurate utilizing a built-in list of commonly misspelled words. Each listener's performance was quantified by calculating the percentage of words correctly transcribed (PWC) in both testing phases, yielding a pretest PWC score and a posttest PWC score for each listener. These PWC scores were used to quantify initial intelligibility (pretest PWC) and intelligibility improvement (posttest PWC after accounting for pretest PWC).

Statistical Analysis.

Initial analyses of the data will include descriptive statistics for the variables of interest, testing for differences in listener cognition between the speaker conditions, testing for average intelligibility improvement between pretest and posttest for the speakers, and calculating correlations between all variables of interest. First, descriptive statistics for listener age and cognitive scores will be calculated for the speaker conditions to ensure that the listeners assigned to each condition are equal. Differences in these variables between the speaker conditions will be tested using the Wilcoxon Rank Sum test since the data did not meet statistical assumptions for parametric testing. Descriptive statistics for pretest and posttest PWC scores for both speaker conditions will also be calculated. Intelligibility improvement between the pretest and posttest will be evaluated for each speaker condition using repeated-measures ANOVA with Cohen's d effect size estimates to test for significant and meaningful intelligibility change due to the perceptual training. Pearson correlation coefficient analyses will be conducted for each speaker condition to assess the correlation between the variables of interest (e.g., cognitive scores, intelligibility).

Elastic net regression models will be used to assess how well the cognitive measures and hearing acuity predict initial intelligibility and intelligibility improvement. Separate elastic net models will be run for each speaker condition as well as for each speaker for initial intelligibility and intelligibility improvement. Therefore, a total of four elastic net models will run during this analysis. Specification of the model includes all main effects as well as all two-way interactions between the variables. The elastic net models will be evaluated using the model accuracy and variable importance metrics. Post hoc assessments will also be conducted to evaluate the nature of the important main effects and interactions. Root mean squared error (RMSE) and R2 will be used to evaluate the predictive accuracy of the cross-validated predictions to incentivize the model to find the predictive relationships that will likely generalize to untrained data. Variable importance will be extracted from each model using permutation. The main effects and interactions selected as important will be assessed via visual analysis of scatterplots depicting the prediction stratified by the variables in the interaction.

All analyses will be completed in the R statistical environment version 4.2.0 or higher.