

Online assessment and enhancement of auditory perception for speech sound errors

Protocol NCT04858022

Document generated 10/21/2020

4.3 Statistical Design and Power: Study 2

This study follows a crossover design in which 40 children with residual speech errors will be allocated to receive 10 weeks/20 sessions of visual-acoustic biofeedback treatment via telepractice followed by a 10-week period of no treatment, or the same two conditions in the reverse order. The primary comparison of interest is the between-group comparison after one group has received treatment and the other has not yet received treatment. The outcome variable will be the proportion of /r/ productions rated as correct, derived from listeners who are blind to treatment condition and timepoint.

All analyses will be conducted following the intent-to-treat principle, with statistician J. Hill generating imputed scores for all cases of attrition. We will analyze perceptual ratings of words elicited in pre-, mid-, and post-treatment probes. For our primary analysis, we will fit a multilevel model with perceptually rated accuracy proportion /r/ correct as the outcome variable. We will test for an effect of phase (pre, mid, or post) and group (treatment-first versus wait-first), as well as the interaction between them. The model will also include random effects for speaker and word, to adjust for the fact that accuracy and treatment response are not uniform across individuals and stimuli.

As an additional measure of functional impact (to be completed as a student project under the REAP mechanism) we will analyze data from a socio-emotional questionnaire [1]. We will use the same statistical model as described above, but with socio-emotional impact score [1] as the outcome variable.

To estimate the likely effect size of a comparison between biofeedback treatment and a no-treatment waitlist condition, we reanalyzed data from a previous single-case experimental study with multiple baselines across participants [90]. In that study, 12 participants were held in a no-treatment baseline condition for an average of 16 sessions. They then received 16 sessions of biofeedback-enhanced treatment, followed by 4 post-treatment maintenance probes. Accuracy of /r/ in word probes was rated by blinded naive listeners recruited online following the same protocols to be used in the proposed research. The effect size (Cohen's d) comparing participants' average performance in the baseline versus post-treatment phase was 1.25. The large size of the effect is not unexpected in the context of a comparison of treatment versus no treatment (as opposed to a comparison between two treatments, which tends to produce a smaller effect size).

Power calculations for the primary analysis were carried out using the PowerUpR package [100] in the R statistical software language [89]. Estimates of the difference in means between groups in Phase 2, while taking account for the multilevel structure of the data, provide strong evidence of power, indicating greater than 80% probability of detecting a between-group effect size of .455. This analysis assumes a small contribution to variance explained based on site and baseline Response Group (10% of variance explained) and an ICC of .25 (based on pilot data). The design has close to 100% power to detect an effect size of the magnitude seen in the pilot data reported above. Our proposed analyses, which will also include data from phases 1 and 3, will have still higher power. Thus, the proposed research is more than adequately powered to detect the hypothesized effect of treatment, even if delivery in the online modality yields a reduction in effect size relative to our estimates from lab-based delivery of treatment.