Correcting Residual Errors With Spectral, Ultrasound, Traditional Speech Therapy (C-RESULTS-RCT)

Protocol NCT03737318

Updated 12/8/2025

MASTER PROTOCOL FOR C-RESULTS RCT

Goal of study: Compare response to two types of visual biofeedback (ultrasound and visual-acoustic) versus traditional motor-based treatment for residual speech errors.
Primary outcome measure: Acoustically measured accuracy of within-session trials from Phase I (Acquisition).
Secondary outcome measure: Perceptually rated accuracy after Phase II (Generalization).
Functional outcome measure: Socioemotional questionnaire before vs after treatment.
Hypothesis: Mixed-effects models investigating the primary outcome measure will show a significant advantage for biofeedback over traditional motor-based treatment as early as the first day of Phase 1, and no later than the final day.
Enrollment: Target 55 kids in each of 2 sites (MSU, Syracuse), total n = 110

0. Pre-screening

- Phone screening to ask about all major exclusionary categories
    - Use Phone Screening Script
    - Additionally, review Parent Questionnaire for any responses that do not meet exclusionary criteria specified in Phone Screening Script.

1. Evaluation Session 1:

- Goal: Determine eligibility for inclusion in study; collect baseline measures for use in predicting response to treatment.
- See "Checklist for C-RESULTS Evaluation - Day 1"
    - Consent and assent; parent history form, demographic and socio-emotional questionnaires
        - Must be between 9;0 and 15;11 at time of enrollment.
        - Must have been exposed to English no later than age 3.
        - English must have equal or dominant status. Balanced bilinguals are okay.
        - Must speak a rhotic dialect of English; presence of one or more non-rhotic speakers in the home is exclusionary.
        - ADHD is admissible; other neurobehavioral disorders (ASD, OCD, Tourette's) are exclusionary.
        - History of CAS / learning disability / dyslexia is admissible if participant meets cutoff scores on eval day 1. Other developmental disorders are exclusionary.
        - History of permanent hearing loss is exclusionary. Temporary hearing loss due to otitis media, including recurrent OM/tubes, is admissible.
    - Hearing screening in soundbooth at 20 dB HL (500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz)—must identify 2/3 tones at 20 dB at each frequency for inclusion.
        - If any items are missed at 20 dB, present the missed item(s) at 25 dB, then re-test at 20 dB. If child is a pass at 20 dB for all frequencies, except ONE ear at ONE frequency that passes at 25 dB, this will still be counted as an overall pass.
        - If non-passing score, refer for full audiological assessment. Participant may enroll if audiological assessment comes back clean.
    - Oral mechanism screening
    - WASI-2 Matrix Reasoning—minimum T score 1.3 SD below age-level mean (T score 37 or higher)
    - CELF-5 Screener—must achieve passing score
        - If failed screener but meets all other criteria, bring child back for Eval Session 2 and administer full CELF-5 Core Language measures; minimum SS 80 for inclusion.
    - Artic test (GFTA-3)—maximum 8th percentile

- Childhood Apraxia Screening
    - Syllable Repetition Task (Shriberg protocol)—must show <4 additions (4 or more additions considered indicative of likely CAS)
    - Linguisystems Articulation Test Inconsistency Screener – must not be marked "inconsistent" on 3 or more items
    - If fail criteria on both, not eligible for the study. If fail criteria on one, then in Evaluation Day 2, administer the Maximum Performance Tasks (Rvachew Protocol) and must score <2 on both apraxia and dysarthria scores.
- Probes for /r/: Sentence, Word, Stimulability
    - Must score no more than 30% on CPP Word probe (50 items). If first clinician rates the child <20%, automatically eligible; if >40%, automatically ineligible. If first clinician rates 20-40% accuracy, second listener required. Two listeners must average to 30% or lower to be eligible.
- Elicit sustained /ɑ/, /i/, /u/, /ɝ/ with CSL

2. Evaluation Session 2:

- Goal: Collect additional sensory measures for use in predicting response to treatment. Administer follow-up measures if non-passing score on the CELF screener or SRT/LAT on Day 1.
- See "Checklist for Evaluation - Day 2"
    - Growth mindset questionnaire
    - Probes for /r/: Sentence, Word, Stimulability
    - CTOPP-2: Elision, Blending, Phoneme Isolation; nonword repetition
    - Rake/wake identification task
    - Stereognosis task
    - If failed CELF Screener, complete CELF-5 Core Language Tests
    - If failed one of SRT or LAT, administer Maximum Performance Tasks (Rvachew protocol)—must score <2 on both dysarthria and apraxia indices

3. Evaluation Session 3

- Goal: Collect additional sensory measures for use in predicting response to treatment.
- See "Checklist for Evaluation - Day 3"
    - Probes for /r/: Sentence, Word, Stimulability
    - Rake-wake discrimination task
    - Bite block task
    - Category goodness task
    - Phonetic awareness task

4. Treatment targets

- Target variants:
    - syllabic /ɝ/, vocalic front vowel context, vocalic back vowel context, consonantal front vowel context, consonantal back vowel context.
- All participants will practice all variants in all phases of treatment.

5. Phase 0: Dynamic assessment

- Dynamic Assessment will consist of a single 90-minute session of traditional motor-based (non-biofeedback) instruction. The goal is to evaluate participants' stimulability in detail and classify them into stimulability categories for randomization.

- Same protocol as Phase I: Acquisition (see below), including both pre-practice and structured practice, but with no biofeedback.
    - Instruction (~25 min): Discuss tongue shapes for /r/. Use "Script_Introducing tongue shapes for r_Dynamic_Assessment_CRESULTS," plus Tiede et al poster, followed by 2-minute break.
    - Pre-practice (until performance criterion met or session-level timer reaches 44 minutes): Relatively unstructured practice guided by the Prepractice Tally Sheet, with model and KP on every trial, followed by a 2-minute break.
    - Structured syllable practice (44 min).
- Based on the treating clinicians' perceptual ratings of participants' performance in Phase 0, participants will be categorized as high responders (>5% accuracy during Dynamic Assessment) or low responders (0-5% accuracy).

6. Randomization

- Overview:
    - Randomization will be supervised by the study statistician at the central site (NYU).
    - Randomization will be stratified by site and severity.
    - Randomization will be proportioned in a 3:3:4 ratio (ultrasound biofeedback : visual-acoustic biofeedback : traditional motor-based treatment) at each site.
- Approach:
    - The study statistician will generate confidential participant treatment assignments in batches of 10, where each batch of 10 corresponds to a combination of site (Montclair State University versus Syracuse University) and response category (high versus low).
    - Random assignments will be generated to ensure the following distribution per batch: 3 individuals assigned to receive visual-acoustic biofeedback, 3 to receive ultrasound biofeedback, and 4 to receive MBT.
    - The first 4 batches will correspond to the first 10 participants recruited in each of these site-by-category combinations. The second 4 batches will correspond to the same site-by-category combinations for the next time period.
    - The PIs will monitor allocation of participants to response categories within each site to understand whether the distribution is reasonably balanced. If it is, the final 4 batches will be assigned similarly. If not, the treatment assignment may be altered.

7. Phase 1: Acquisition

- Overview:
    - Goal: Measure response to a brief but intensive period of treatment (3 sessions, each 1.5 hours long, in 1 week).
    - Use document "Checklist for CRESULTS Intensive Treatment—Traditional," "Checklist for CRESULTS Intensive Treatment—Ultrasound," or "Checklist for CRESULTS Intensive Treatment—VA"
- Pre-practice (max 44 min; could be less if participant meets performance criterion to advance to structured practice):
    - In first session only, provide condition-specific instructions (approximately 25 minutes):
    - Using "Script_Introducing ultrasound_CRESULTS" document, "Script_Introducing VA_CRESULTS," or "Script_Introducing traditional treatment_CRESULTS." After first session, pre-practice should be broken into two 20-min blocks of practice separated by one obligatory 2-minute break.
    - Pre-practice should involve relatively unstructured, highly interactive elicitation; provide models and placement cues using document "Articulator placement cues_CRESULTS."

- Keep track of participant's correct productions using document "Pre-Practice Tally Sheet." Pre-practice will start by eliciting /ɝ/. A model and KP feedback should be provided on every trial. Once a particular target has been produced three times in a fully correct fashion (use a strict standard) in pre-practice, move on to another target. (Can shape from the previous target.)
    - If participant produces at least five syllable targets three times in a fully correct fashion (=15 total), move on to Structured Practice.
- If participant is assigned to a biofeedback condition, visual biofeedback remains available throughout pre-practice.
- Take a 2 minute break (obligatory) between pre-practice and structured practice.

- Structured practice (44 min; could be more or less depending on performance and time spent in pre-practice):
    - Number of trials
        - Elicit up to 200 trials in structured practice. Target a minimum of 150 trials, but if cumulative session duration reaches 90 minutes, terminate session.
        - After 100 trials, CPP will prompt user to take a 2-minute break. CPP should not resume function until 2 minutes have elapsed. If CPP prompts a break with 5 minutes or less remaining in the session, OK to override the break (press s).
    - Targets
        - To limit variability of practice, only 5 different syllables will be selected for practice in Acquisition. Syllables will be /ɝ/ plus one randomly selected syllable for the other four variants. The same set of syllables will be kept constant across sessions.
        - Practice will occur in blocks of 10 consecutive trials on the same syllable (e.g., 10 /ra/), after which a new syllable will be addressed (e.g., 10 /re/--but note that in the fully blocked condition, the same syllable should occur in two consecutive blocks of 10). CPP should cycle through the set of target syllables, randomizing order of presentation.
    - Scoring and feedback:
        - Clinician scores each production as correct (1) or incorrect (0).
            - For the first 50 trials (i.e., the first block of each syllable), clinician should be cued to provide KP feedback after 5/10 trials, randomly selected from within the biofeedback trials. KP feedback must make reference to both the visual display and to articulator placement; it must also terminate with a direct model for the client to imitate.
            - For the remaining trials, clinician should be cued to provide KP feedback on 3 trials, randomly selected from within the biofeedback trials. KR feedback should be provided on 2 trials other trials, randomly selected. KR feedback can be verbalized but does not have to be (i.e., text presented by CPP alone is sufficient).
        - Cue biofeedback for the first 8/10 trials in each block.
        - In each block of 10 trials, the 9th trial will be flagged for acoustic measurement: a pure tone will sound before the trial, flagging the trial for acoustic measurement via forced alignment. The same trial will be elicited with **no verbal or visual feedback**.
    - The CPP will provide reinforcing messages at the end of each block.
        - If the participant is advancing to a new level, the CPP will report this.
        - Otherwise, CPP should report number of trials completed.
        - Milestone/number of trials message should be paired with a message to encourage continued effort (Growth Mindset).

8. Midterm administration of safety check

- At the start of Generalization Session 7: Follow Protocol for Safety Check
- Collect participant responses to the following questions:
    - "Since the beginning of this study, have you experienced any possible side-effects?"
    - "Is there anything you would like us to know about how the study may have affected you, positively or negatively?"

9. Phase II: Generalization

- **At the start of the generalization sessions 1 and 9:** Re-administer Stimulability Probe, CPP word- and sentence-level /r/ probes.
- All participants will receive two treatment sessions per week for 8 weeks (16 sessions). Sessions will be a maximum of 60 minutes in duration.
- Use CPP and Checklist for Tx Generalization CRESULTS
- Each session begins with pre-practice (maximum 10 minutes).
    - Pre-practice should include two stimulus items from each of the 5 target variants and focus on the level of complexity the child left off at in the previous session, plus or minus one level.
        - Use CPP examiner view to identify two target items per variant (10 total) that will be targeted in pre-practice; write them in the pre-practice tally sheet.
        - Terminate pre-practice once child has produced all 10 items three times in a fully correct fashion, or 10 minutes has elapsed (whichever comes first).
- Structured practice:
    - Number of trials
        - Maximum 250 /r/ trials. Target minimum 200 trials, but if cumulative session duration reaches 1 hour, terminate session.
        - After 100 trials, CPP will prompt user to take a 2-minute break. CPP should not resume function until 2 minutes have elapsed. If CPP prompts a break with 5 minutes or less remaining in the session, OK to override the break (press s).
    - Targets
        - Syllables/words targeted in each session will be randomly selected by the CPP program from standard lists of words/syllables. A separate set of syllables/words should be selected at the start of each session.
        - Words will be elicited via the CPP program in blocks of 10 trials.
        - At the base level, sessions should feature 10 items total, with 2 syllables/words representing each variant. However, if the participant moves up a level (see *Across-session manipulations*), a new set of words will be drawn, resulting in >10 words occurring in the session.
    - Within-session manipulations: The following changes may take place within a treatment session.
        - In each block of 10 trials, the 9th trial will be flagged for acoustic measurement: a pure tone will sound before the trial, cueing the clinician to avoid talking over the child and flagging the trial for acoustic measurement via forced alignment. The same trial will be elicited with **no verbal or visual feedback**.
        - At the start of a session, the participant's starting point in the challenge point hierarchy is based on their performance in the last session.
        - Feedback frequency

- First level: 2 KP+model trials, randomly scheduled from within biofeedback trials. 3 KR trials, randomly scheduled.
- Next level: 1 KP+model trial, randomly scheduled from within biofeedback trials. 4 KR trials, randomly scheduled.
- Next level: 1 KP+model trial, randomly scheduled from within biofeedback trials. 3 KR trials, randomly scheduled.
- Next level: 1 KP+model trial, randomly scheduled from within the first 8 trials (no biofeedback at this level). 2 KR trials, randomly scheduled.
- Final level: 1 KP+model trial, randomly scheduled from within the first 8 trials (no biofeedback at this level). 2 KR trials, randomly scheduled, but with self-rating. Once clinician enters rating, instead of displaying clinician's rating, display prompt: "What did you think?" CPP then accepts a second keypress from the clinician representing the participant's response (1 or 0).
- **Note our formula for KP (artic + display + model):**
    i. Reference what the child is doing or should be doing with the articulators
    ii. Refer to the visual display IF in a biofeedback condition
    iii. End with an explicit verbal model (yes, even if the child has advanced to random practice. At the sentence level, model only the target word).
- **Biofeedback feedback frequency is reduced at the same time as the first four changes in KP/KR feedback frequency.** Levels: 80%-50%-20%-0%.
    i. Visual icon cues each trial as biofeedback on/off.

- Prosodic manipulation, i.e. stimulus ends with ./!/? (3 levels: none, blocked, random)
- Word shape/context (9 levels: syllables, 1-syllable simple words, 1-syllable words with competing phoneme, multisyllabic simple words, multisyllabic words with competing phoneme, words in a carrier phrase, words in sentences, words in sentences with multiple /r/ words, independently generate a sentence with a randomly selected target word).
    - At the carrier phrase level and above, words will be randomly selected from any of the preceding levels (**excluding** the syllable level) and presented in the appropriate context by the CPP.

- In blocked practice, each block begins with a direct model for the client to imitate.
    - Fade models once random practice is initiated.
    - If client makes a reading error, try to cue the target without modeling—start with a semantic cue, then a rhyme cue.
- Across-session manipulations: If the participant's accuracy across all trials in the previous session was equal to or greater than 80%, the CPF program will adjust the order of presentation.
    - 3 levels: Fully blocked (each target word presented in two consecutive blocks [2 words per variant; program randomly rotates through all variants]); random blocked (each block elicits 10 trials of a single word, but across blocks, different words and variants can be presented in random order [4 words per variant]); fully random (different words and variants represented within each block).

- The CPP will provide reinforcing messages at the end of each block.
    - If the participant has achieved a new milestone (e.g., advancing to a new level; first time getting 6/10 correct in a block), the CPP will report this.
    - If there is no new milestone, CPP should report number of trials completed.

- Milestone/number of trials message should be paired with a message to encourage continued effort (Growth Mindset).

10. Post-test I (immediately after treatment)

- Goal: Evaluate progress; re-administer a subset of sensory measures elicited initially.
- See "Checklist for C-RESULTS_Evaluation_PostTX_RCT"
- Probes for /r/ and other sounds: Stimulability Probe for C-RESULTS, CPP word- and sentence-level /r/ probes, CPP tongue complexity probe
- Rake-wake identification task
- Rake-wake discrimination task
- Goodness judgment task
- Phonetic awareness task
- **Final administration of safety check**
    - Follow Safety Monitoring Protocol
    - Collect participant responses to the following questions:
        - "Since the beginning of this study, have you experienced any possible side-effects?"
        - "Is there anything you would like us to know about how the study may have affected you, positively or negatively?"
    - Also collect Post-Treatment Social-Emotional Impact and Therapy Experience questionnaires

11. Post-test II (6 weeks after treatment)

- Goal: Evaluate change in time since treatment; administer perturbation tasks (NYU/MSU only) and accompanying perception tasks (all sites).
- See 1-Checklist for CRESULTS_Evaluation_PostTx2_RCT_ALLTasks (NYU/MSU) or 1-Checklist for CRESULTS_Evaluation_PostTx2_RCT_NoPerturbation (Syracuse) in CRESULTS Planning Folder/4c - Evaluation_COVID_CRESULTS/Post-Tx Evaluation 2 - RCT
- Probes for /r/ (if a child is being evaluated more than 6 weeks post treatment, don't administer probes; use 1-Checklist for CRESULTS_Evaluation_PostTx2_RCT_NoProbes)
- Head-had identification task
- Head-had discrimination task
- Reflexive perturbation task (NYU/MSU only)
- Adaptive perturbation task (NYU/MSU only)

12. Analysis Plan

- All analyses will be conducted following the intent-to-treat principle. For participants who drop out, missing data will be multiply imputed based on information collected about them prior to their loss to follow-up.
- For our primary outcome (Acquisition), we will fit a multilevel model with measures nested within study participants and words. The dependent variable will be F3-F2 distance in /r/ targets produced in each day of Phase 1 treatment. The primary independent variable of interest is group (traditional motor-based treatment versus biofeedback, with visual-acoustic and ultrasound biofeedback types pooled for this analysis). The model will also adjust for site, performance response category based on Dynamic Assessment (high responder versus low responder), acquisition day (1, 2, or 3, treated as categorical), and pre-treatment accuracy (mean F3-F2 distance across /r/ sounds in the word probe elicited in the first pre-treatment evaluation). If a within-session time trend is supported by the data, we will extend the model to account for time trends across trials within each day of treatment. Random intercepts will be included to reflect the fact

that observations are nested within speakers and words, and random slopes that improve overall model fit (for speaker, acquisition day; for word, performance response category and pre-treatment accuracy) will be retained following model comparison using likelihood ratio tests. Likelihood ratio tests will also be used to assess the significance of fixed effects and interactions in the final model. Likelihood ratio tests will also be used to test the hypothesis of an advantage for biofeedback over MBT. Tests will be performed in two different model specifications. In one, the overall effect across days will be tested by focusing on the coefficient on the treatment variable. The other specification will interact the treatment variable with time in order to test whether the effect varies across days, since the hypothesized advantage for biofeedback over MBT could emerge as early as the first day or as late as the third day of Phase 1 (Acquisition). Although a positive effect of biofeedback is hypothesized, two-sided hypothesis tests will be used to be conservative.

- For our secondary outcome (Generalization), we will use a linear mixed-effects model to examine word probe data collected in one immediate pre-treatment and one immediate post-treatment session. The outcome variable, "p̂-correct," will reflect the proportion of naïve listeners who rated the /ɹ/ sound in a given production as correct. Independent variables will include time point (pre- versus post-treatment) and treatment group (biofeedback versus MBT), as well as the interaction between them. As for our primary measure, the model will adjust for site, performance response category based on Dynamic Assessment, and pre-treatment accuracy. Random intercepts for speaker, item, and listener will be included. Random slopes of response category and pre-treatment accuracy on word will be included only as warranted by model comparison. As above, likelihood ratio tests will be used to assess the significance of fixed effects and interactions in the final model. Again, while a positive effect of biofeedback is hypothesized, we will use a two-sided hypothesis test to be conservative. For this second model, we will additionally test for the presence of a significant time-by-condition interaction in which the magnitude of improvement associated with biofeedback significantly is allowed to vary with time.
- Finally, we will analyze survey data using a similar model, with impact score as the outcome variable and fixed effects of time, treatment group, and the time-treatment interaction, as well as site, performance response category based on Dynamic Assessment, and pre-treatment accuracy. We will again estimate the interaction between time and treatment group to evaluate whether the functional impact of treatment differs across biofeedback versus MBT conditions.

13. Fidelity Checking Plan

- Treatment sites perform fidelity checks on each other to avoid breaking blinding for data site.
- Fidelity checks should be conducted on an ongoing basis.
- Randomly select a subset of sessions from each session type:
    - For DA, complete fidelity checks on 20% of subjects per site.
    - For Acquisition, randomly select one of the three sessions per subject per site.
    - For Generalization, randomly select one from the first half and one from the second half of treatment per subject per site.
- Re-score a subset of trials to check agreement across sites.
    - Re-score tone trials only in acquisition sessions slated for fidelity checking.

Data analysis plan

All analyses will be conducted following the intent-to-treat principle, comparing participants based on the treatment assigned regardless of subsequent exposure. To maintain the full sample, missing data will be multiply imputed for participants who drop out of the study using the *mice* package [45] in the R statistical software language [46]. This approach will capitalize on information collected on these participants prior to their loss to follow-up.

To assess impact on our primary outcome, the acoustic accuracy measure of F3-F2 distance, we will fit a multilevel model using Phase 1 (Acquisition) data. The primary effect of interest is a comparison between MBT and biofeedback treatments, with visual-acoustic and ultrasound biofeedback types pooled for this analysis. This initial model will also adjust for site, performance response category (high responder versus low responder based on a Dynamic Assessment), Acquisition day (1, 2, or 3), and pre-treatment accuracy (mean F3-F2 distance across /ɹ/ sounds in the word probe elicited in the first pre-treatment evaluation). If a within-session time trend is supported by the data, we will extend the model to account for time trends across trials within each day of treatment. Random intercepts will be included to reflect the fact that observations are nested within speakers and words, and random slopes that improve overall model fit (for speaker, acquisition day; for word, performance response category and pre-treatment accuracy) will be retained following model comparison using likelihood ratio tests.

Likelihood ratio tests will also be used to test the hypothesis of an advantage for biofeedback over MBT. Tests will be performed in two different model specifications. In one, the overall effect across days will be tested by focusing on the coefficient on the treatment variable. The other specification will interact the treatment variable with time in order to test whether the effect varies across days, since the hypothesized advantage for biofeedback over MBT could emerge as early as the first day or as late as the third day of Phase 1 (Acquisition). Although a positive effect of biofeedback is hypothesized, two-sided hypothesis tests will be used to be conservative.

Our second analysis will fit a linear mixed-effects model using word probe data collected in one immediate pre-treatment and one immediate post-treatment session. The outcome variable will reflect the proportion of naïve listeners who rated the /ɹ/ sound in a given production as correct. The model will include an indicator for time point (pre- versus post-treatment) and for treatment group (biofeedback versus MBT), as well as the interaction between them. As was the case for our primary measure, the model will adjust for site, performance response category based on Dynamic Assessment, and pre-treatment accuracy. Random intercepts for speaker, item (the 50 target words making up the probe measure), and listener will be included. Random slopes of response category and pre-treatment accuracy on word will be examined and included only as warranted by model comparison. As above, likelihood ratio tests will be used to assess the significance of the overall treatment effect. Again, while a positive effect of biofeedback is hypothesized, we will use a two-sided hypothesis test to be conservative.

For this second model, we will additionally test for the presence of a significant time-by-condition interaction in which the magnitude of improvement associated with biofeedback significantly is allowed to vary with time. If we observe a difference in our primary measure (Acquisition) but not our secondary measure (Generalization), this will be interpreted as evidence that biofeedback is more efficient, but not more effective, than MBT.

The final analysis will evaluate changes in impact scores from the 11-item survey assessing the social, emotional, and academic consequences of RSE from pre- to post-treatment time points. A mixed model similar to that used in the previous analyses will be fit including time (pre- versus post-treatment), treatment group (biofeedback versus MBT), and the time-treatment interaction as the primary predictors. The model will also adjust for site, performance response category based on Dynamic Assessment, and pre-treatment accuracy. We will again estimate the interaction between time and treatment group to evaluate whether the functional impact of treatment differs across biofeedback versus MBT conditions.