Official Title of the study: Effects of Altered Auditory Feedback on Speech Fluency

NCT Number: [NCT ID not yet assigned]

Date: 05/01/2023

Data Collected:

- Audio (.wav format)
    - For each subject:

		Testing Sessions					
		Screening	Baseline	Week 1	Week 2	Week 3	Week 4
Condition	Baseline (No Auditory Feedback)	Open response clips (90s) x 3 Passage reading clip (up to 4 min) x 1	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2 Passage reading clip (up to 4 min) x 1
	Unaltered		Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2
	Whisper		Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2
	Reverb		Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2
	Harmonize		Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2
	Final (No Auditory Feedback)		Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2	Open response clips (90s) x 2

- Survey data (Multiple choice answers and typed answers), containing biographical information and quantitative + qualitative insights into the experiences of participants while using the app, and overall
    - Initial Survey
    - optional Overall Assessment of the Speaker's Experience of Stuttering (OASES™)
    - Final Survey
- Usage data from Mumble Melody application
    - Length of use for each mode (milliseconds)
    - Time of use for each mode (Year, Month, Date, H:M:S)

Data Processing:

1. **Generate Transcript:** Whisper[1] software will be used to generate a first draft of text from audio recordings
2. **Manual Marking of Stuttering-Like Disfluencies:** Two data coders will manually mark the scripts to characterize Stuttering-like-disfluencies (SLD), including blocks, repetitions, and prolongations, similar to the coding used in the Stuttering Severity Instrument-4
    - Data coders will have received training on identifying SLDs
    - Data coders will be blind to when the data were acquired
3. **SLD counts:** For each condition, the SLD count will be used to establish a count of disfluencies per syllable (SLD/syl) (as a percentage), per minute (SLD/m), and per word (SLD/w) for each condition (Baseline, Unaltered Mode, Reverb Mode, Harmonize Mode, Final).
    a. **SLD/syl:** The syllable counts and SLD/syllable values will be calculated using an online tool such as the CLAN (Computerized Language ANalysis)[2] software.

Data Analysis:

General methods which will be used for analyses:

1. **Calculation of ΔF_{Mode}**: The percentage of fluency variation (ΔF_{Mode}) between Condition A and Condition B will be calculated based on the equation below:

$$\Delta F_{Mode} = \frac{[SLD/syl]_{ConditionA} - [SLD/syl]_{ConditionB}}{[SLD/syl]_{ConditionA}}$$

- The value of the ΔF_{Mode} will be a percentage change in the number of disfluencies. A positive ΔF_{Mode} will indicate increased fluency for Condition B compared to Condition A and a negative ΔF_{Mode} will indicate decreased fluency for Condition B compared to Condition A.
    - ΔF_{Mode}= 50% indicates the disfluencies have been halved in Condition B compared to Condition A
    - ΔF_{Mode}= 0% indicates no change in fluency
    - ΔF_{Mode}= 100% indicates no disfluencies in Condition B

2. **Statistical Analysis:**
    - **Statistical Significance:** Wilcoxon signed-rank tests will be used to on pairs of SLD/syl values to assess the statistical significance of fluency improvement between conditions across subjects
    - **Significance Threshold:** $\alpha = 0.05$

Calculation of ΔF_{Mode}:

For each session (Baseline Session, Week 1, Week 2, Week 3, Week 4)

1. Calculate ΔF_{Mode},percentage improvement in dis/syl for each subject for each mode
    a. Condition A: Baseline Condition - No Auditory Feedback
    b. Condition B: Each mode condition (Unaltered, Whisper, Reverb, Harmonize, Final - No Auditory Feedback)

Primary Outcome:
Significant improvement in fluency in one of four modes (Unaltered, Whisper, Reverb, Harmonize) between the baseline testing session and the end of week 4.
Modes = Unaltered, Whisper, Reverb, Harmonize
For EACH of these modes, we will evaluate:
Improvement in Testing Session Compared to Each Other: Week 4 Compared to Baseline Testing Session

1. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing SLD/syl for Week 4 with the Baseline Testing Session. Results in 1 p-value
2. Step 2: If the p-value < α:
    a. Then we reject H_0 = that Week 4 does not have statistically greater SLD/syl, as compared to the Baseline Testing Session, for mode X

Secondary Outcomes:
Significant improvement in fluency in one of four modes (Unaltered, Whisper, Reverb, Harmonize) between the baseline testing session and end of week 1, 2, and 3.
Modes = Unaltered, Whisper, Reverb, Harmonize
For EACH of these modes, we will evaluate:
Improvement in Testing Session Compared to Each Other: Week 1,2, or 3 Compared to Baseline Testing Session

1. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing SLD/syl for Week 1,2 or 3 with the Baseline Testing Session. Results in 1 p-value
2. Step 2: If the p-value < α:
    a. Then we reject H_0 = that Week 1,2, or 3 does not have statistically greater SLD/syl, as compared to the Baseline Testing Session, for mode X

Significant improvement in fluency in the Baseline (No Auditory Feedback) condition between baseline and end of week 1, 2, 3, and 4.
For the Baseline (No Auditory Feedback) condition, we will evaluate:
Improvement in Testing Session Compared to Each Other: Week 1,2, 3, or 4 Compared to Baseline Testing Session

1. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing SLD/syl for Week 1,2 3, or 4 with the Baseline Testing Session. Results in 1 p-value
2. Step 2: If the p-value < α:
    a. Then we reject H_0 = that Week 1,2, 3, or 4 does not have statistically greater SLD/syl, as compared to the Baseline Testing Session, for the Baseline (No Auditory Feedback) condition

Significant improvement in fluency in the Final (No Auditory Feedback) condition between baseline and end of week 1, 2, 3, and 4.
For the Final (No Auditory Feedback) condition, we will evaluate:
Improvement in Testing Session Compared to Each Other: Week 1,2, 3, or 4 Compared to Baseline Testing Session

3. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing SLD/syl for Week 1,2 3, or 4 with the Baseline Testing Session. Results in 1 p-value
4. Step 2: If the p-value < α:
    a. Then we reject H_0 = that Week 1,2, 3, or 4 does not have statistically greater SLD/syl, as compared to the Baseline Testing Session, for the Final (No Auditory Feedback) condition

Significant improvement in fluency for the Final (No Auditory Feedback) condition in comparison to the baseline (No Auditory Feedback) condition, for each testing session:
For each session (Baseline Session, Week 1, Week 2, Week 3, Week 4):
SLD/syl Comparison: (A) Baseline Condition - No Auditory Feedback (B) Final Condition - No Auditory Feedback

1. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing the final condition with the baseline condition. Results in 1 p-value
2. Step 2: If the p-value < α:
    a. Then we reject H_0 = that the final condition does not have statistically greater SLD/syl, as compared to the baseline (no auditory feedback)

Significant improvement in fluency for each mode in comparison to the baseline (No Auditory Feedback) condition, for each testing session
For each session (Baseline Session, Week 1, Week 2, Week 3, Week 4):
SLD/syl Comparison: (A) Baseline Condition - No Auditory Feedback **(B)** Each Mode (Unaltered, Whisper, Reverb, Harmonize)

1. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing each mode with the baseline condition. Results in 4 p-values across subjects.
2. Step 2: If the p-value < α:
    a. Then we reject H_0 = that each mode does not have statistically greater SLD/syl, as compared to the baseline (no auditory feedback)

Significant improvement in fluency for each non- control mode (Whisper, Reverb, Harmonize) in comparison to the control mode (Unaltered), for each testing session
For each session (Baseline Session, Week 1, Week 2, Week 3, Week 4):
SLD/syl Comparison: (A) Unaltered Mode (B) Whisper, Reverb, Harmonize Modes

1. Step 1: Conduct Wilcoxon one-sided signed-rank test comparing each non-control mode with the control (Unaltered) mode. Results in 3 p-values across subjects.
2. Step 2: If the p-value < α:
    a. Then, for each non-control mode, we reject H_0 = that each non-control mode does not have statistically greater SLD/syl, as compared to the control (unaltered) mode

References:

[1] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust speech recognition via large-scale weak supervision.” OpenAI. Accessed: Oct. 12, 2022. [Online].
[2] MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.