**Acoustic Measures Versus Crowdsourced Listeners’ Ratings: Evaluating The Efficacy of Treatment for Hypokinetic Dysarthria**

Tara McAllister Byun¹, Christopher Nightingale¹, Michelle Swartz², Lorraine Olson Ramig³,⁴
New York University¹, Montclair State University², University of Colorado at Boulder³, Columbia University⁴

---

**Introduction**
- Individuals with Parkinson’s Disease (PD) often develop voice and speech deficits termed hypokinetic dysarthria.
- Characteristics include reduced vocal loudness, abnormal voice quality, reduced prosodic variation, and decreased magnitude and precision of articulator movements.
- **Lee Silverman Voice Treatment** (LSVT) is an intensive one-month regimen that trains patients to recalibrate to using a louder voice and a higher level of effort during speech.
- Three randomized controlled trials have reported significant changes in the acoustics of PD patients’ speech subsequent to LSVT treatment. [1,2,3]
  - Specifically, the ratio between the second formant frequency in /i/ vs /u/ (F2i/F2u) is used as an index of magnitude of articulator movement.
  - However, evaluation of the perceptual impact of these changes has been limited and has used expert listeners.
- To know whether treatment will impact functional communication, it would be ideal to use naïve listeners’ ratings. [4]
- Technologies for online crowdsourcing allow rapid collection of ratings from large samples of nonexpert listeners.
  - Amazon Mechanical Turk (AMT) is a crowdsourcing platform where anyone can post an electronic task or sign up to complete tasks for pay.
  - Crowdsourcing could enable a shift toward a more naturalistic standard to evaluate speech interventions. [4,5]

---

**Objectives**
- Determine whether crowdsourced listener ratings can be used to replicate findings from a published study documenting the efficacy of LSVT.
- **Sapir et al.** [6] found significant changes in expert listeners’ ratings and acoustic measures (F2i/F2u) in vowels produced before vs after LSVT therapy.
- **Hypothesis:** Responses aggregated across naïve listeners recruited online will show a significant difference between pre- and post-therapy tokens.

---

**Methodology**
- Phrases spoken by 14 PD patients with moderate-severe hypokinetic dysarthria were acquired by Sapir et al. [6]
  - Target words “key” and “stew” were isolated and normalized in intensity.
  - 4 PRE and 4 POST words for each participant were paired; total of 56 pairs.
  - AMT was used to recruit 35 naïve listeners who were self-reported native speakers of English and had no history of speech/language impairment.
  - Online experiment presentation software [7] was used to present pairs of tokens in random order.
  - Raters selected the production that was most clearly articulated.

---

**Results**
- Post-treatment tokens received a higher percentage of “more clear” ratings than pre-treatment tokens.
  - Mixed-effects logistic regression revealed that this difference was significant ($\beta=1.24, SE=.46, p<.01$).
  - A significant time-word interaction indicated that the rating difference was greater in “stew” than “key.”
  - At the individual level, there was a fair amount of variability in magnitude of pre-post change in perceptual rating.
  - There was generally good agreement between the acoustic and perceptual data.
  - In Figure 2, subjects are ordered according to the magnitude of acoustic change (pre-post difference in F2i/F2u ratio).
  - Across subjects, the acoustic measure of individual change was significantly correlated with the magnitude of change in the percentage of tokens rated “more clear” ($R = .62, p = .01$).

---

**Conclusions**
- These results support the original conclusion of Sapir et al. [6]: participants made significant gains in speech clarity from pre- to post-treatment phases of the study.
  - These gains are evident not only in acoustic measures and trained listeners’ ratings [6], but also in judgments by non-expert listeners recruited online.
  - These results support the validity of crowdsourcing as a means to obtain ratings of disordered speech data.
- Clinical researchers are encouraged to consider online crowdsourcing as an efficient means to obtain naturalistic judgments of treatment efficacy.

---