Use of a Telehealth Platform to Automatically Assess Prosodic Contours in Parkinson Disease

Andrew H. Exner\textsuperscript{1}, Vikram Ramanarayanan\textsuperscript{2}, David Pautler\textsuperscript{2}, Sandy Snyder\textsuperscript{1}, Hardik Kothare\textsuperscript{2}, Jackson Liscombe\textsuperscript{2}, Oliver Roesler\textsuperscript{2}, William Burke\textsuperscript{2}, Michael Neumann\textsuperscript{2}, David Suendermann-Oeft\textsuperscript{2}, & Jessica E. Huber\textsuperscript{1}

\textsuperscript{1}Purdue University, \textsuperscript{2}Modality.AI
Relevant Disclosures

• Financial Disclosures
  • Modality.AI, Inc., Salary & Shares
    • Vikram Ramanarayanan – Chief Scientific Officer
    • David Pautler – Founder, Chief Technology Officer
    • Hardik Kothare – Research Scientist
    • Jackson Liscombe – Research Scientist
    • Oliver Roesler – Research Scientist
    • William Burke – Director
    • Michael Neumann – Research Scientist
    • David Suendermann-Oeft – Founder, Chief Executive Officer
  • Purdue University, Salary
    • Andrew Exner – doctoral student, trainee on NIH T32 grant
    • Sandy Snyder – Research Associate
    • Jessica Huber – Professor

• Non-Financial Disclosures
  • Jessica Huber – Medical advisory board for Rock Steady Boxing
Parkinson Disease

• Neurodegenerative disorder primarily affecting motor system
  • Motor, cognitive, and sensory
• Speech: Hypokinetic Dysarthria
  • Hypophonia
  • Variable rate
  • Breathiness
  • Monopitch
  • Monoloudness
  • Imprecise consonants
  • Decreased intelligibility
Hypokinetic Dysarthria: Prosody

• Hypokinetic prosody abnormalities include
  • Reduced $f_0$ range and variability (monopitch)
  • Reduced ability to use prosody for emotional expression
  • Not everything is impaired
    • Lexical stress is spared
    • Need: what is impaired vs what is not and why?
    • i.e., is sentence mode differentiation impaired?
Rationale

• Needs
  • Accessibility of care by people with PD
  • Need for monitoring of dynamic symptoms
  • Burden of technical assessment and measurement

• Solution
  • Conversational artificial intelligence agent
  • Automatic computation and delivery of relevant patient data

• Features
  • Automated, customizable assessment
  • Convenient time
  • Home environment
  • Minimal technological requirements
  • Automatic computation of speech acoustic metrics, facial kinematic metrics, and limb motor function
  • User-friendly dashboard for healthcare providers
  • Symptom tracking over time
Aims and Hypotheses

• Compare automatic measures produced by the Modality system with default Praat settings and data extraction algorithms to human-generated measurements calculated by members of the Purdue Motor Speech Lab in order ascertain the feasibility and reliability of automated analytics for assessing the prosody of people with PD.

• Hypothesis: There will be no significant differences between the automated $f_0$ measures generated through default Praat settings and those made by human researchers.
Methods: Participants

• $n = 40$ people with PD; 23 age- and sex-matched controls

• Inclusion criteria:
  • Age 30-85
  • Dx idiopathic PD
  • Internet access
  • Device w/ microphone & camera
  • Self-reported adequate hearing and vision
  • Fluency in English

• Exclusion criteria:
  • Dx neurological disease other than PD
  • Hx HNC cancer or surgery (except for implantation of DBS)
  • Hx voice disorder or pulmonary disease
  • Recent Hx smoking (<5 years)
  • More than moderate cognitive impairment <10 on MoCA)
Methods: Initial Visit

• WebEx meeting with lab staff member
  • Discuss Study
  • Obtain Consent
  • Obtain Medical History
  • Complete Montreal Cognitive Assessment
  • Orientation to System Access
  • Receive individualized link to complete online assessments
Methods: Conversations with Tina

- Number of Assessments: 4
- Frequency of Assessments: 1/week
  - Median 8 days, Mean 10 days
- Timing: When convenient for participants, on-state of PD medication
- All tasks completed each session
- Total Duration: 15-20 minutes

- Speech Tasks
  - Sustained vowels
  - Sentence Intelligibility Test (SIT)
  - Reading 1 paragraph of Rainbow Passage
  - Short narrative
  - *Intonational prosody*
  - Monologue

- Non-Speech Tasks
  - Abbreviated oral mechanism exam
  - Finger tapping

- Surveys
  - Parkinson Disease Questionnaire (PDQ-39)
  - Communication Participation Bank, Short (CPIB-S)
  - Task Load Index (TLX)
Methods: Intonational Prosody Task

• Participants presented with a short scenario and asked to say the sentence provided.
  • Five pairs of sentences (three words each)
  • Same except for the prosodic falling or rising contour cued by different scenarios

• Examples

  • Tina: “You just got back from holidays in Florida. Jane asks if the weather was nice. Now you say...”
  • Target: “It was hot.” (Statement)

  • Tina: “Jane says her vacation to Alaska was too hot. Now you say...”
  • Target: “It was hot?” (Question)
Methods: Measurements

- **Key measurements of intonational contour direction and variability**
  - Minimum $f_0$ (Hz)
  - Maximum $f_0$ (Hz)
  - Standard deviation of $f_0$ (Hz)
  - Range of $f_0$ (maximum - minimum $f_0$) (Hz)

- **Human-Corrected Measurements**: standard Praat settings to assess pitch points within the Manipulation file
  - Deleting pitch points during voiceless segments
  - Adding pitch points not identified by Praat (e.g., rapid pitch changes, occurring above/below Praat’s default)
  - Correcting pitch points during diplophonia

- **Modality.AI system (1)**: automatic extraction of the same $f_0$ values using Praat’s default settings
  - No alteration of the default pitch contour extracted

- **Modality.AI system (2)**: optimized $f_0$ extraction with optimized parameters based on subset of data
Methods: Pitch Correction

- High prevalence of aperiodic voicing, periodic vocal fry, and diplophonia
  - Of 788 utterances, 486 (61.7%) contained aperiodic voicing
  - Of these, the mean percentage of aperiodic voicing per utterance was 13.3% (± 9.2 % SD)
Example: Noise marked with pitch periods during a /t/
Example: No $f_0$ marked during falling vocalization
Example: Correction of falling contour
Example: Roughness/Fry (Female) (Original)
Example: Roughness/Fry (Female) (Corrected)
Methods: Statistical Analysis

• To determine whether the automated measurements differed significantly from the clinician-researcher measurements
  • ICC estimates and their 95% confidence intervals were calculated
  • Excel
  • Single-rating, absolute-agreement, two-way random-effects model with one rater across all subjects
Results: Means and Standard Deviations of Fundamental Frequency Measures in a Prosody-Specific Speech Task, Human-Corrected vs Unoptimized Automated (n = 40 PD, 23 controls)

<table>
<thead>
<tr>
<th>Frequency Measure</th>
<th>Human-Corrected</th>
<th>Automated</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum f0 (Hz)</td>
<td>132.7</td>
<td>123.48</td>
<td>0.611</td>
</tr>
<tr>
<td>Maximum f0 (Hz)</td>
<td>268.55</td>
<td>334.27</td>
<td>0.410</td>
</tr>
</tbody>
</table>

ICC: Intraclass Correlation Coefficient
Results: Means and Standard Deviations of Fundamental Frequency Variation in a Prosody-Specific Speech Task, Human-Corrected vs Unoptimized Automated ($n = 40$ PD, $23$ controls)

- Standard Deviation of Fundamental Frequency (f0SD) Hz
  - Human-Corrected: 39.74
  - Automated: 56.81
  - ICC: 0.419

- Range of Fundamental Frequency (Max - Min) (Hz)
  - Human-Corrected: 135.85
  - Automated: 210.79
  - ICC: 0.331
Results: Mean Absolute Errors of Fundamental Frequency Metrics

<table>
<thead>
<tr>
<th></th>
<th>Default</th>
<th>Optimized (Combined)</th>
<th>Optimized (Predicted Sex)</th>
<th>Optimized (Known Sex)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>13.4</td>
<td>8.97</td>
<td>8.06</td>
<td>7.08</td>
</tr>
<tr>
<td>SD</td>
<td>19.6</td>
<td>7.8</td>
<td>6.16</td>
<td>6.05</td>
</tr>
<tr>
<td>Minimum</td>
<td>15.69</td>
<td>14.83</td>
<td>13.7</td>
<td>12.88</td>
</tr>
<tr>
<td>Mean Absolute Error (Hz)</td>
<td>69.15</td>
<td>16.98</td>
<td>13.79</td>
<td>12.71</td>
</tr>
</tbody>
</table>
Results: Means and Standard Deviations of Fundamental Frequency Measures in a Prosody-Specific Speech Task, Human-Corrected vs Optimized Automated ($n = 40$ PD, 23 controls)

![Bar chart showing minimum and maximum fundamental frequency (f0) measures with ICC values.]

- Minimum $f_0$ (Hz): Human-Corrected 132.7, Automated 139.35, ICC: 0.691
- Maximum $f_0$ (Hz): Human-Corrected 268.55, Automated 265.48, ICC: 0.848
Results: Means and Standard Deviations of Fundamental Frequency Variation in a Prosody-Specific Speech Task, Human-Corrected vs Optimized Automated ($n = 40$ PD, 23 controls)

![Graph showing Standard Deviation of Fundamental Frequency (Hz)](image1)

![Graph showing Range of Fundamental Frequency (Max - Min) (Hz)](image2)

- **Standard Deviation of Fundamental Frequency (f0SD)**
  - Human-Corrected: 39.74 Hz
  - Automated: 36.61 Hz
  - ICC: 0.763

- **Range of Fundamental Frequency (Max - Min) (Hz)**
  - Human-Corrected: 135.85 Hz
  - Automated: 126.13 Hz
  - ICC: 0.758
Discussion: Reliability of Prosodic Measures

• Initial substantial differences in human-corrected and automated measures of all 4 $f_0$ measures
  • Minimum $f_0$ differences were small, likely of little to no clinical significance
  • Other differences were larger and likely of clinical significance
• Following optimization, differences are significantly reduced
• Remaining Issues to be Addressed
  • How to detect and correct prevalent aperiodic voicing
  • How to prevent autocorrelation method from assigning pitch periods to unvoiced segments without changing pitch floor/ceiling
Discussion: Clinical Feasibility

• Patients can perform this task independently over the internet
• System can identify pitch periods with moderate-to-good accuracy
• System reported intonation measures have moderate-to-good reliability with human-corrected measures
Future Directions

• Determine whether optimized parameters can generalize to a larger sample of people w/ PD (in process)

• Nuclear tone analysis (in process)
  • Compare whole-utterance intonation contour
  • To nuclear tone (final word in utterance) contour
  • To determine which is a better representation of speaker’s intonation (for PD)

• Compare objective measurements to subjective ratings of rate and naturalness (in process)
  • E.g., PDQ-39, CPIB-S, clinician ratings of speech severity
Acknowledgments

• Undergraduate researchers in the Purdue Motor Speech Lab
• NIH T32 Training Grant (SLHS @ Purdue, 2021-2022)
• Modality.AI Team
Questions?
References


Example: Roughness (Male) (Original)
Example: Roughness (Male) (Corrected)
Example: Roughness/Fry (Female) (Original)
Example: Roughness/Fry (Female) (Corrected)
Results: Means and Standard Deviations of Fundamental Frequency Measures in a Prosody-Specific Speech Task, Human-Corrected vs Unoptimized Automated ($n = 40$ PD, 23 controls)

<table>
<thead>
<tr>
<th></th>
<th>Human-Corrected Mean (SD)</th>
<th>Automated Mean (SD)</th>
<th>ICC (95% CI) (agreement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum $f_0$ (Hz)</td>
<td>132.70 (36.03)</td>
<td>123.48 (37.42)</td>
<td>0.611 (0.538, 0.672) (moderate)</td>
</tr>
<tr>
<td>Maximum $f_0$ (Hz)</td>
<td>268.55 (94.32)</td>
<td>334.27 (134.41)</td>
<td>0.410 (0.218, 0.551) (poor)</td>
</tr>
<tr>
<td>$f_0$SD (Hz)</td>
<td>39.74 (25.92)</td>
<td>56.81 (37.08)</td>
<td>0.419 (0.243, 0.550) (poor)</td>
</tr>
<tr>
<td>$f_0$ Range (Hz)</td>
<td>135.85 (79.51)</td>
<td>210.79 (131.98)</td>
<td>0.331 (0.122, 0.488) (poor)</td>
</tr>
</tbody>
</table>
Results: Means and Standard Deviations of Fundamental Frequency Measures in a Prosody-Specific Speech Task, Human-Corrected vs Optimized Automated \((n = 40 \text{ PD, 23 controls})\)

<table>
<thead>
<tr>
<th></th>
<th>Human-Corrected Mean (SD)</th>
<th>Automated Mean (SD)</th>
<th>ICC (95% CI) (agreement)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum (f_0) (Hz)</strong></td>
<td>132.70 (36.03)</td>
<td>139.35 (34.56)</td>
<td>0.691 (0.637, 0.735) (moderate)</td>
</tr>
<tr>
<td><strong>Maximum (f_0) (Hz)</strong></td>
<td>268.55 (94.32)</td>
<td>265.48 (87.60)</td>
<td>0.848 (0.827, 0.866) (good)</td>
</tr>
<tr>
<td>(f_0)SD (Hz)</td>
<td>39.74 (25.92)</td>
<td>36.61 (22.72)</td>
<td>0.763 (0.727, 0.794) (moderate)</td>
</tr>
<tr>
<td>(f_0) Range (Hz)</td>
<td>135.85 (79.51)</td>
<td>126.13 (74.64)</td>
<td>0.758 (0.722, 0.790) (moderate)</td>
</tr>
</tbody>
</table>
F0 Tuning Process

• Single measurer identified 575 turns of interest
• Ran Praat's default pitch calculation algorithm
  • "Sound: To Pitch", autocorrelation method, time step = 0.0, pitch floor = 75Hz, ceiling = 600Hz
• Human-corrected contours
• Compared f0 metrics (e.g., mean) from the reference contours and our baseline predicted ones using mean absolute error (MAE). **Green bars.**
• Ran 7,186 pitch calculations using different Praat algorithms and settings.
• Extracted f0 metrics and found Praat settings that minimize MAE. **Blue bars.**
• Identified the optimal settings for known sex-based cohorts. **Red bars.**
• Ran a machine learning experiment to predict patient sex. **Yellow bars.**
• Implemented code for doing the last experiment.
Tuning Observations

• It is well known that automated pitch extraction is best when one uses sex-specific settings.

• This was shown in our results as well, though the increase in metric accuracy was not that big (compare red to green bars).

• Nevertheless, using a machine classifier to predict sex (since we may not always know it at the time of metric extraction) was almost as good as knowing the sex a priori (yellow vs green bars) and still better than a sex-agnostic pitch extract algorithm (red vs yellow bars).

• All three algorithms, however, show drastic improvement over the default Praat settings, with most reduction in error being for max F0.
Tuning Parameter Space

- \( f0\_type = "ac" \) (autocorrelation) or \( "cc" \) (cross-correlation)
- \( \text{pitch\_floor} = 10-600 \text{ Hz} \)
- \( \text{pitch\_ceiling} = 100-700 \text{ Hz} \)
- \( \text{max\_candidates} = 1-100 \)
- \( \text{very\_accurate} = "on" \) or \( "off" \)
- \( \text{silence\_thresh} = 0.01-1.0 \)
- \( \text{voicing\_thresh} = 0.1-1.0 \)
- \( \text{octave\_cost} = 0.0-1.0 \)
- \( \text{octave\_jump\_cost} = 0.1-1.0 \)
- \( \text{voiced\_unvoiced\_cost} = 0.1-1.0 \)
- \( \text{kill\_octave\_jump} = "yes" \) or \( "no" \)
- \( \text{smoothing} = 0 \) - 100
Old Praat F0 Code

• To Pitch: 0, 75, 600
New Praat F0 Code: Step 1 Predict Sex

- `f0_type$ = "cc"`
- `pitch_floor = 55.0`
- `max_candidates = 9`
- `very_accurate$ = "off"`
- `silence_thresh = 0.07`
- `voicing_thresh = 0.49`
- `octave_cost = 0.03`
- `octave_jump_cost = 0.5`
- `voiced_unvoiced_cost = 0.16`
- `pitch_ceiling = 350.0`
- `kill_octave_jump$ = "no"`
- `smoothing = 25`

To Pitch (cc): 0.01, pitch_floor, max_candidates, very_accurate$, silence_thresh, voicing_thresh, octave_cost, octave_jump_cost, voiced_unvoiced_cost, pitch_ceiling

Smooth: smoothing

- `sex$ = "F"`
- `if (mean_f0 <= 156.67555)`
  - `sex$ = "M"`
- `elif (mean_f0 <= 189.532093) and (min_f0 >= 90.901626) and (min_f0 <= 127.179596)`
  - `sex$ = "M"`
- `endif`
New Praat F0 Code: Step 2 Create Sex-Optimized Contour

```
ll_type$ = "ac"
very_accurate$ = "off"
kill_octave_jump$ = "no"
if sex$ == "F"
  pitch_floor = 125.0
  max_candidates = 3
  silence_thresh = 0.03
  voicing_thresh = 0.57
  octave_cost = 0.02
  octave_jump_cost = 0.6
  voiced_unvoiced_cost = 0.2
  pitch_ceiling = 500.0
  smoothing = 24
else
  pitch_floor = 75.0
  max_candidates = 14
  silence_thresh = 0.04
  voicing_thresh = 0.56
  octave_cost = 0.01
  octave_jump_cost = 0.45
  voiced_unvoiced_cost = 0.19
  pitch_ceiling = 350.0
  smoothing = 22
endif
To Pitch (ac): 0.01, pitch_floor, max_candidates, very_accurate$, silence_thresh, voicing_thresh, octave_cost, octave_jump_cost, voiced_unvoiced_cost, pitch_ceiling
```