THE IPHONE AS A USEFUL AND VALID TOOL IN LARYNGOLOGY

Jeremy Hornibrook,¹ Emily Lin,² & Tika Ormond²

¹Department of Otolaryngology, Christchurch Hospital
²Department of Communication Disorders, University of Canterbury
Christchurch, New Zealand
Outline

- Background
  - iPhone
  - Acoustic measures
- Research question
- Normal voice (Cross-system validation)
- Pathological voice:
  - Methodology
  - Results
    - Normals vs. patients
    - Case presentations (pre and post-surgery)
- Conclusions
Background - iPhone

- iPhone (first released in: 2007): a multi-media enabled mobile phone with advanced
  - Computing capabilities
  - Connectivity: Internet and email access
- iPhone 3G & up: sampling rate = 48,000 Hz (Lossless)
Background – Acoustic Measures

- **Time-based measures:**
  - **Fundamental frequency (F0):** affected by mass and stiffness
    - edema (smokers): decreased F0 (Sorensen & Horii, 1982)
    - Voice patients have difficulties maintaining a constant pitch (Kotby, Titze, Saleh, & Berry, 1993).
    - Speaking F0 changes after treatment of functional voice (Roy & Taskco, 1994)
  - **Perturbation measures**
    - Jitter (or percent jitter; %Jit): cycle-to-cycle pitch variation
      (e.g., Eskenazi, Childers, & Hicks, 1990; Dejonckere, Remacle, Fresnel-Ebaz, Woisard, Crevier-Buchman, & Millet, 1996; Wolfe & Martin, 1997; Bhuta, Patrick, & Garnett, 2004)
    - Shimmer (or percent shimmer, %Shim): cycle-to-cycle amplitude variation
      (e.g., Dejonckere et al., 1996; Wolfe & Martin, 1997; Bhuta et al., 2004)
    - Signal to noise ratio (SNR): energy ratio between periodic & aperiodic components
      (e.g., Wolfe & Martin, 1997; Brockmann, Storck, Carding, & Drinnan, 2008)

Less hoarse = decreased jitter & shimmer; increased SNR
Background – Acoustic Measures

- **Frequency-based measures (i.e., spectral measures):**
  - **Sentence:** Spectral tilt (ST): amplitude difference between the highest spectral peak between 0-1 kHz and that between 1-5 kHz
    - Higher ST = vocal hypofunction (Löfqvist, 1987; Mendoza, Munoz, & Valencia Naranjo, 1996)
  - **Vowel:**
    - Formants One and Two frequencies (F1 and F2): Affected by vocal tract configuration (constriction or tongue placement);
      - Larger vowel space area (F1-F2 plot of vowels) = greater intelligibility
        - (e.g., Bradlow, Toretta, & Pisoni, 1996; Roy, Nissen, Dromey, & Sapir, 2009; Turner, Tjaden, & Weismer, 1995; Weismer, Jeng, Laures, Kent, & Kent, 2001)
    - **H1-H2:** amplitude difference between the first two harmonics
      - Smaller H1-H2 = less breathy or thicker voice
        - (e.g., Klatt & Klatt, 1990; Hillenbrand, Cleveland, & Erickson, 1994; de Krom, 1995; Hillenbrand & Houde, 1996; Stone, Cleveland, Sundberg, & Prokop, 2003)
    - **Singing power ratio (SPR):** amplitude difference between the highest spectral peak between 0-2 kHz and that between 2-4kHz
      - Smaller SPR = greater voice projection power
        - (e.g., Omori, Kacker, Carroll, Riley, & Blaugrund, 1996)
Research Question

- Are iPhone recordings adequate for acoustic assessment of speech and voice quality in:
  - Normal voice
  - Pathological voice, for the purpose of:
    - Identifying voice aberrations: i.e., differentiating normal from pathological voices
    - Monitoring voice changes: e.g., detecting pre and post-treatment differences?
Normal Voice

- Poster presentation, January 12:
  Lin, E. & Hornibrook, J. “The Suitability of iPhone Recordings for the Acoustic Measures of Speech and Voice Quality.”

- Methodology:
  - Simultaneous voice recordings (sustained vowels & sentences) using an iPhone & a direct digitization system
  - 11 normal speakers (6 females & 5 males), aged 27-67 years (Mean = 41.8, SD = 16.7)
  - Acoustic measures:
    - Sentence-based: ST
    - Vowel (50 ms segment): F0, perturbation (%Jit, %Shim, SNR), H1-H2, SPR, F1, & F2

- Findings:
  - Relatively high cross-system correlations \( r = 0.74 \) to \( 0.98 \), demonstrating adequate parallel validity.
  - However, mean normalized absolute inter-system differences are optimal (i.e., lower than 20%) only for F0, F1, & F2, suggesting that most quality-related acoustic measures are not directly comparable.
Pathological Voice

- **Methodology:**
  - Participants: 22 patients (10 males & 12 females; aged 25-92 yrs, Mean = 54.8, SD = 18.5), including 5 pre & post-surgery cases
  - Participant’s task: read the first 6 sentences of “Rainbow passage”
  - Measures:
    - **Subjective assessment:** 2 listeners, with 3rd sentence randomly presented twice & judged on a 4-point (0-normal, 1-slight, 2-moderate, 3-severe) scale GRBAS (G: grade of hoarseness, R: roughness, B: breathiness, A: asthenia, S: strained)
    - **Acoustic measures:** same as previous study; using TF32 (Milenkovic, 2000)
      - Sentence-based measure: ST
      - Vowel-based measures (/i/, /a/): F0, %Jit, %Shim, SNR, H1-H2, SPR, F1, & F2
      - Vowel space area: F1-F2 plot of /i/, /a/, /u/
  - Analysis:
    - **Group comparisons:**
      - Normal vs. patient (for males & females separately)
      - Pre vs. post-surgery (for 5 patients):
        - Visual analysis: Normalized scores (z scores), calculated via a linear transformation using the means and standard deviations (SD) of the normal data in previous study, i.e.,
          \[ z = \frac{\text{raw score} - \text{Mean}}{\text{SD}} \]
Measurement Reliability

- **Subjective assessment:** the two ratings of the same token for all data (11 normals + 22 patients + 5 post-surgery cases) were highly correlated except for roughness (good intra-judge but poor inter-judge reliability).

- **Acoustic measures:**
  - Automatic computer derivations of acoustic measures: 100% reliability, except for errors due to variation in vowel segmentation.
  - All data from the five pre and post-surgery cases (vowel-based measures: 5 patients X 2 visits X 2 vowels) were re-segmented and analyzed. The measure-remeasure reliabilities were found to be relatively high for all vowel-based measures:

<table>
<thead>
<tr>
<th>Subjective GRBAS Ratings</th>
<th>Intra-judge</th>
<th>Inter-judge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>r1</td>
</tr>
<tr>
<td>G</td>
<td>38</td>
<td>0.91</td>
</tr>
<tr>
<td>R</td>
<td>38</td>
<td>0.88</td>
</tr>
<tr>
<td>B</td>
<td>38</td>
<td>0.82</td>
</tr>
<tr>
<td>A</td>
<td>38</td>
<td>0.91</td>
</tr>
<tr>
<td>S</td>
<td>38</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acoustic Measures</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0</td>
<td>20</td>
<td>0.99</td>
</tr>
<tr>
<td>%Jit</td>
<td>20</td>
<td>0.99</td>
</tr>
<tr>
<td>%Shim</td>
<td>20</td>
<td>0.98</td>
</tr>
<tr>
<td>SNR</td>
<td>20</td>
<td>0.97</td>
</tr>
<tr>
<td>F1</td>
<td>20</td>
<td>0.93</td>
</tr>
<tr>
<td>F2</td>
<td>20</td>
<td>0.88</td>
</tr>
<tr>
<td>H1-H2</td>
<td>20</td>
<td>0.97</td>
</tr>
<tr>
<td>SPR</td>
<td>20</td>
<td>0.96</td>
</tr>
</tbody>
</table>
Normals vs. Patients

- **Spectral Tilt (ST):** Patients > Normals

Result of Rank Sum tests:
- **Male:** Patients (n = 10) have a **significantly higher median ST** than normals (n = 5).
- **Female:** Patients (n = 12) has a higher median ST than normals (n = 6) but the difference is not statistically significant.

* : Significant at 0.05 level
Normals vs. Patients

- \( \% \text{Jit} \): Patients > Normals

* : Significant at 0.05 level
Normals vs. Patients

- **%Shim**: Patients > Normals

* Significant at 0.05 level
Normals vs. Patients

- **SNR:** Patients < Normals

* **SNR (in dB)**

---

* Significant at 0.05 level
Normals vs. Patients

- **Subjective (GRBAS):** Patients > Normals

*: Significant at 0.05 level
Case One: Left Vocal Fold Paralysis

- Male, aged 79 yrs (M79), aortic arch aneurysm

PRE

POST (medialization laryngoplasty)
Case One: M79, VF Paralysis & Medialization

Improvement after surgery:
- lower pitch: decreased F0 (/a/)
- less hoarse:
  - decreased %Jit & %Shim
  - increased SNR
- less breathy: decreased H1-H2
- more intelligible:
  - Increased vowel space area
  - F2: increased for /i/, decreased for /a/
Case Two: Polyp on LVF

- Male, aged 33 yrs (M33)
Case Two: M33, Polyp & Microsurgery

Improvement after surgery:
- less hoarse:
  - decreased %Jit (for /i/) & %Shim
  - increased SNR (for /i/)
- more breathy (or thinner voice): increased H1-H2
- more intelligible:
  - Increased vowel space area
    - F1: increased for /i/ & /a/
    - F2: increased for /i/ & /a/

---

**Acoustic Parameter**

**/i/**

**/a/**

**z score**
Case Three: Inflammatory myoblastic tumor

- Male, aged 34 yrs

PRE

POST (microsurgery)
Case Three: M34, IM tumor & Microsurgery

Improvement after surgery:
- lower pitch: decreased F0
- less hoarse:
  - decreased %Jit & %Shim
  - increased SNR
- more intelligible:
  - Increased vowel space area
    - F1: decreased for /i/, increased for /a/
    - F2: increased for /i/ & /a/
Case Four: Mass lesion on RVF

- Male, aged 46 yrs

PRE

POST (microsurgery)
Case Four: M46, Lesion & Microsurgery

**Acoustic Parameter**

/i/:
- PRE
- POST

/a/:
- PRE
- POST

**Improvement after surgery:**
- lower pitch: decreased F0
- less hoarse:
  - decreased %Jit & %Shim for /i/
  - increased SNR
- more intelligible:
  - Increased vowel space area
    - F2: increased for /i/, decreased for /a/
Case Five: Nodules

- Female, aged 39 yrs

PRE

POST (microsurgery)
Case Five: F39, Nodules & Microsurgery

**Improvement after surgery:**
- lower pitch: decreased F0
- less hoarse:
  - decreased %Jit & %Shim
  - increased SNR
- less breathy: decreased H1-H2
- more intelligible:
  - F1: decreased for /i/, increased for /a/
Spectral Tilt: Pre vs. Post

- Case 3 (M34, IM tumor): ST decreased after treatment (i.e., less hypofunction)
- Case 5 (F39, nodule): ST increased after treatment
Summary of Main Findings

- Relatively high measure-remeasure and acceptable between-system reliabilities for all acoustic measures included in this study.

- Most measures are adequate for detecting voice improvement, especially:
  - Perturbation measures (%Jit, %Shim, SNR) & vowel space area consistently demonstrate positive changes after treatment.

- Frequency-based measures are more variable:
  - H1-H2 may reflect subtle changes in breathiness (or voice thickness).
  - Changes in SPR may be vowel-dependent.
  - Changes in ST may be gender or pathology-dependent.
Conclusions

- Voice recordings using iPhone are adequate for acoustic measurement of speech and voice quality.
- However, due to large inter-subject variations, most of these measures are more useful for intra-subject comparison (to monitor changes within individuals) than for norm-referenced comparison.
References


References - continued


References - continued


