Physiologic Voice Assessment

Auditory-Perceptual, Acoustic, Aerodynamic,
CAPE-V

The blue spot is on the key again
How hard did he hit him
We were away a year ago
We eat eggs every Easter
My mama makes lemon muffins
Peter will keep at the peak
Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

Name: ___________________________ Date: ____________

The following parameters of voice quality will be rated upon completion of the following tasks:

1. Sustained vowels, /a/ and /i/ for 3-5 seconds duration each.

2. Sentences production:
   - a. The blue spot is on the key again.
   - b. How hard did he hit it?
   - c. We were away a year ago.
   - d. We eat eggs every Easter.
   - e. My name makes lemon maffins.
   - f. Peter will keep at the peak.

3. Spontaneous speech in response to: "Tell me about your voice problem." or "Tell me how your voice is functioning."

Legend:
- \( C = \) Consistent
- \( I = \) Intermittent
- \( M = \) Mildly Deviant
- \( M = \) Moderately Deviant
- \( S = \) Severely Deviant

<table>
<thead>
<tr>
<th>Overall Severity</th>
<th>M</th>
<th>I</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roughness</td>
<td>M</td>
<td>I</td>
<td>S</td>
</tr>
<tr>
<td>Breathiness</td>
<td>M</td>
<td>I</td>
<td>S</td>
</tr>
<tr>
<td>Strain</td>
<td>M</td>
<td>I</td>
<td>S</td>
</tr>
<tr>
<td>Pitch (Indicate the nature of the abnormality)</td>
<td>M</td>
<td>I</td>
<td>S</td>
</tr>
<tr>
<td>Loudness (Indicate the nature of the abnormality)</td>
<td>M</td>
<td>I</td>
<td>S</td>
</tr>
</tbody>
</table>

COMMENTS ABOUT RESONANCE: NORMAL OTHER (Provide description):

ADDITIONAL FEATURES (for example, dysphonia, fry, huskiness, asthenia, aphonia, pitch instability, trueness, waxy/gurgly, or other relevant terms):

Clinician: ___________________________
Overall severity

Roughness

Breathiness

Strain

Pitch/Loudness
Auditory-perceptual assessment

Listener judgment is the gold standard

Types of rating scales

- Categorical, ordinal
- Equal-appearing interval (EAI)
- Visual analog scale (VAS)
- Direct magnitude estimation (DME)

Issues

- Comparing stimuli to internal scale versus external exemplar
- Individual perceptual habits and biases
Acoustic Analysis

Measuring what we hear

Quantifying acoustic integrity of voice

Outcome measurement
Selected from steady-state portion of sustained “ah”
• F0 – mean fundamental frequency
• Jitter – cycle-to-cycle variation in frequency
• Shimmer – cycle-to-cycle variation in intensity
• NHR – spectral-based ratio: energy between the harmonics divided by the harmonics

Voice signal typing is important
Selected from steady-state portion of sustained “ah”

- **F0** – mean fundamental frequency
- **Jitter** – cycle-to-cycle variation in frequency
- **Relative average perturbation (RAP)** – average absolute difference between a period and the average of it and its two neighbors, divided by the average period
- **Shimmer** – cycle-to-cycle variation in intensity
- **NHR** – spectral-based ratio: energy between the harmonics

Voice signal typing is important.

I’m hoarse when I talk.
Selected from steady-state portion of sustained “ah”

- F0 – mean fundamental frequency
- Jitter – cycle-to-cycle variation in frequency
- Relative average perturbation (RAP) - average absolute difference between a period and the average of it and its two neighbors, divided by the average period
- Shimmer – cycle-to-cycle variation in intensity
- NHR – spectral-based ratio: energy between the harmonics divided by the harmonics

Voice signal typing is important
Signal typing

- **Type I**
  - Nearly periodic waveforms that don’t have any qualitative changes during the time interval of analysis
  - Any modulations or subharmonic frequencies that may be present have an energy at least one order of magnitude less than that of F0

- **Type II**
  - Signals contain sudden qualitative changes in the interval to be analyzed, or that have modulating or subharmonic frequencies whose energy is comparable to that of the F0

- **Type III**
  - No apparent periodicity; thus, not valid for F0 measurement
  - Signals may be chaotic
• Can’t use in connected speech
• Not valid for severe/aperiodic voices
• Don’t change with treatment
• OR
• Change with time
Acoustic analysis of running speech

Cepstral analysis

- The cepstrum is a Fourier transform of a Fourier transform
- The cepstrum isolates F0 while preserving and revealing characteristics of the filter

Better ecological validity because running speech is analyzed

Can be used on samples with unsteady pitch and loudness
### Cepstral analysis - measures

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| **Cepstral peak prominence (and SD)** | • Relative amplitude of the dominant cepstral peak  
• Correlates with severity of breathiness (Hillenbrand et al., 1994)  
• Increased abnormality associated with decrease in overall CPP (Wolfe et al., 2000) |
| **Cepstrum-derived HNR** | • Strong single predictors of severity for both rough and breathy voice samples (de Krom, 1995)  
• Strong multiple regression predictors of overall severity (de Krom, 1995) |
| **Pitch sigma (and SD)** | • Derived from F0 standard deviation  
• Measure of long-term instability in which variations in frequency occur more slowly than the glottal vibration itself (Hartelius et al., 1997)  
• Predictive of breathiness and hoarseness |

Shaheen Awan, PhD
FIGURE 1. Cepstral analysis using smoothing algorithm and linear regression analysis from Hillenbrand software.
Cepstral Spectral Index of Dysphonia

\[ CSID_{cv} = 148.68 - (5.91 \times CPP) - (11.71 \times \sigma_{CPP}) - (1.31 \times LHRatio) - (3.09 \times \sigma_{LHRatio}) \]

• A multifactorial estimate of voice severity.
• Correlates with the labeled visual analog scale for overall severity (in %) used in the CAPE-V.
• Discriminate normal from disordered voices.
• Sensitive to treatment change.


Awan SN, Roy N, Jette ME, Meltzner GS, Hillman RE. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clin Linguist Phon.* Sep 2010;24(9):742-758.

Typical adult female

CSID score = -13.9 (low is good).
Mild adult female

CSID = 36.82, mildly hoarse, roughness
Correlation between acoustic and auditory perceptual analyses of voice across 4 common voice disorders

Amanda I Gillespie, PhD
Ali Lewandowski, MA
Jackie Gartner-Schmidt, PhD
University of Pittsburgh Voice Center, Pittsburgh, PA

Shaheen N Awan, PhD
Bloomsburg University, Bloomsburg, PA
Outcomes

Hypothesis-driven outcomes testing
Study Objectives

1. To determine if auditory perceptual ratings of voice severity correlate with acoustic measurements.

2. To determine if that correlation is dependent on voice disorder or voice stimulus.
Methods

- 40 participants
  - 10 each with: MTD, lesions, atrophy, UVFP
- Strict single diagnosis/treatment
- SLP/Laryngologist team
## Methods

Data collected before and after treatment

<table>
<thead>
<tr>
<th>MTD</th>
<th>Lesions</th>
<th>Atrophy</th>
<th>UVFP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td>Voice therapy</td>
<td>Excision</td>
<td>Injection augmentation</td>
</tr>
<tr>
<td><strong>f/up</strong></td>
<td>6 mo post-tx</td>
<td>12 mo post-tx</td>
<td>6 mo post-tx</td>
</tr>
</tbody>
</table>
Methods

How hard did he hit him
We were away a year ago
We eat eggs every Easter
Peter will keep at the peak
/a/
All stimuli together
Methods

Acoustic analysis
- CSID
- 2 raters

Auditory-perceptual analysis
- 10 blinded raters
- CAPE-V overall severity
Results

Reliability

• Auditory perceptual
• 7/10 raters $r > .80$
"How hard did he hit him?"
"We were away a year ago."
"We eat eggs every Easter."
"Peter will keep at the peak."

CSID vs Auditory Perceptual Rating

CSID
Auditory Perceptual Rating
## Disorder-Specific Results

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTD</td>
<td>• NS</td>
</tr>
</tbody>
</table>
| Lesions  | • good correlations for all stimuli  
          • NS differences pre-post: “we were away a year ago”, /a/ , all stimuli combined. |
| Atrophy  | • little pre-post treatment difference  
          • NS differences pre-post: “we were away a year ago”, “Peter will keep at the peak”, /a/ |
| UVFP     | • good correlations for all stimuli  
          • NS differences pre-post: “we were away a year ago”, all stimuli combined. |
Discussion

• Correlation between acoustic (CSID) and auditory-perceptual ratings of overall voice severity as a function of voice stimulus and voice disorder.

• “We were away a year ago”, all stimuli combined: best correlation between acoustics and perception.
Discussion

• CSID tended to overestimate voice severity as compared to the perceptual ratings.
  • Ex: “how hard did he hit him”
  post-treatment CSID = 33.13
  perceptual rating = 21.27
Discussion

• Weaker correlations following treatment.
• MTD: correlations between CSID and perceptual ratings were all non-significant.
  • Vocal variability
  • Strain may have a limited effect on the CPP, resulting in reduced correspondence between the CSID and perceptual ratings.
Take home

Acoustic and Auditory-Perceptual assessment of voice in natural, connected speech, is superior to assessment of single sounds.
Video of acoustic collection