Clinical Evaluation in Head and Neck Cancer: Voice Quality and Speech Intelligibility Outcomes

Philip C. Doyle, Ph.D., CCC-SLP
Voice Production and Perception Laboratory
Department of Otolaryngology Head and Neck Surgery
Schulich School of Medicine and Dentistry
University of Western Ontario
London, Ontario, Canada
**The Focus of Today’s Presentation**

*Voice Quality (VQ) and Speech Intelligibility (SI)*

- Two inextricably linked elements of one’s communication capacity and/or ability
- One is relatively well-defined (SI), one is not (VQ)
- SI assessed with scaling and identification
- Evidence suggests DME procedures are ideal for SI (prothetic, a magnitude or quantity)
- VQ and SI evaluation remains challenging
“Speech [is] abnormal when it deviates so far from the speech of other people that it calls attention to itself, interferes with communication, or causes the speaker or his listeners to be distressed”

Charles Van Riper, 1978, p. 43
A Clinical Problem

The appropriate method for measuring what listeners hear when they listen to voices remains an unresolved issue, and providing accurate, replicable, valid measures of vocal quality presents significant challenges.

From: Kreiman, Vanlancker-Sidtis, & Gerratt (2004)

Also applied to SI in those with head and neck cancer!
A Basic Concern: Listener Experience

Experienced Listeners:
- Increased exposure/experience with dysphonic VQs, cannot guarantee satisfactory levels of interjudge agreement (Kent, 1996).
- Experienced listeners may present with professional biases
  - (Kreiman & Gerratt 1993).
  -

Naïve Listeners:
- Perceptual judgements likely based on a uniform standard that seeks comparison to normal (Kreiman & Gerratt 1993).
Clinical Considerations

- Oncologic safety
- Survival
- Secondary complications
- Speech
- Swallowing
- General physical health
- Overall quality of life (QoL)
“Cut to the Chase”

- Multiple dimensions of voice and speech may be scaled using visual analogue (VA) methods
- Reliability appears reasonable
- Naïve listeners preferred, but professionals are acceptable
- Time efficient and economical
- Intelligibility evaluation may be scaled, but integration with direct measures ideal
Voice Quality defined:

- That component of speech which gives the primary distinction to a given speaker’s voice when pitch and loudness are excluded.
- Involves both phonatory and resonance characteristics (e.g., in dysphonic voice—harshness, breathiness, etc.)

From: [http://www.biology-online.org/dictionary/voice](http://www.biology-online.org/dictionary/voice)

A composite auditory-perceptual judgment of the “vocal” signal
How has VQ been Measured?

Use of scaled methods – equal appearing interval (EAI) scales most common in literature

EAI scales: fixed, pre-defined scale values suggesting “equality” of perceptual distance between ratings
Speech Intelligibility defined:

- The information a listener can recover from the speech signal (Kent, 1993)
- The accuracy with which a normal-hearing listener can understand the spoken signal. This level of accuracy may be assessed in numerous ways (isolated sounds, words, etc.)
How is SI Measured?

- Overall SI determined by scaling – a global evaluation; again often using EAI

  Highly <<<           >>>>Highly
  Intelligible       Unintelligible

- Detailed SI determined by direct identification of spoken stimuli
Rating Scales

Direct Magnitude Estimation (DME): a method of perceptual ratio scaling
Modulus = 100
Twice as “pleasant” = 200 Half as “pleasant” = 50

Equal-Appearing Interval (EAI) Scales: fixed, pre-defined scale values suggesting “equality” of perceptual distance between ratings

Unpleasant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Very Pleasant
Why doesn’t EAI work?
Preliminary Considerations

- Is there an overlap between *speech intelligibility* (SI) and *voice quality* (VQ) in those with head and neck cancer?

- From a strict measurement perspective, the answer is unclear.

- However, in those with H&NCa, a listener cannot always divide her/his attention between the two!
Relationships among dimensions?

<table>
<thead>
<tr>
<th></th>
<th>SEV</th>
<th>NAT</th>
<th>ACC</th>
<th>PL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEV</td>
<td>...</td>
<td>.930**</td>
<td>.931**</td>
<td>.945**</td>
</tr>
<tr>
<td>NAT</td>
<td>...</td>
<td>...</td>
<td>.883**</td>
<td>.911**</td>
</tr>
<tr>
<td>ACC</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>.943**</td>
</tr>
</tbody>
</table>
Selecting SI Stimuli

- Selection of stimuli ideally conforms to client’s potential limitations
- What are the anticipated speech production deficits? For example...
  
  Laryngeal cancer – change in source
  Oropharyngeal cancer – change in filter (articulation and resonance), vocal tract transmission, and potentially source deficits
EL pleasantness and acceptability both curvilinear. They cannot be scaled with EAI methods.

From: Beaudin, Eadie, & Doyle (2005)
Some Suggestions: Speech Intelligibility

- For alaryngeal speech, no accepted standard exists

- For other H&NCA groups, it may be best to utilize concepts and principles from the dysarthrias

- Weiss and Basili (1985) – EL speech

- Kent, Kent, & Rosenbek (1987) – 19 acoustic-phonetic contrasts

- Phoneme vs. word scoring
How do listeners’ make judgments?

Regardless of intelligibility or alterations in voice quality, listeners consistent order speaker “proficiency” based on global indices of pleasantness, acceptability, naturalness, and effectiveness.

(Chalmers & Doyle, 1994; Doyle, 1999 Niemi & Doyle, 1999; Hare & Doyle, 2000)
Proposed VQ Rating Scale

Based on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

The parameter voice quality will be rated for the following sentence:

“The rainbow is a division of white light into many beautiful colors.”

Legend:  MI = Mildly Deviant  MO = Moderately Deviant  SE = Severely Deviant

<table>
<thead>
<tr>
<th>Overall Severity</th>
<th>MI</th>
<th>MO</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listener Comfort</td>
<td>MI</td>
<td>MO</td>
<td>SE</td>
</tr>
</tbody>
</table>

SCORE /100
Auditory-Perceptual Judgments Defined:

Those judgments, ratings, perceptions that quantify or qualify features, dimensions, attributes, or parameters, either partial or complete, with the goal of evaluating speaker “performance”.

What is known?

For TE speakers:
  Naturalness and Pleasantness are linear
  Severity and Acceptability are curvilinear
  *Eadie & Doyle, 2002a,b*

For EL speakers:
  Pleasantness and Acceptability are curvilinear
  *Beaudin, Eadie, & Doyle, 2005*
A Fundamental Problem

- All individuals treated for head and neck cancers are indeed individuals – expectation of specific outcome is problematic
- Treatments differ
- Negative consequences of treatment differ
- Treatment outcomes are not uniform across many domains of functioning
Some definitions:

- **Voice severity:**
  “a comprehensive measure of how ‘good’ or ‘poor’ the voice sample is judged to be by the listener”.

- **Pleasantness:**
  “how ‘pleasant’ you find the speaker’s voice as a listener. The rating should be made exclusive of speech intelligibility, dialectal variation, and/or accent”.
Clinical Implications

When assessing psychophysical character of non-normal vocal signals, evaluation of appropriate method of scaling mandatory.

When evaluating non-normal, unusual, or “bizarre” voices, careful consideration of method of perceptual analysis is critical.
VA-EAI Scores Plotted as Function of Geometric Means of DME Scores
LC-EAI Scores Plotted as Function of Geometric Means of DME Scores
PC vs. VA scaling – 3 Speaker Groups

Listener Comfort

Speaker

PC
VAS

HE1 HE3 HE5 HE7 TE1 TE3 TE5 TE7 R1 R3 R5 R7
Overall Voice Quality

Although TE speech approximates normal $f_0$, meets or exceeds intensity, and mirrors normal durational characteristics, voice quality remains non-normal!

Rough, noisy, and aperiodic
Statement of Problem

Although physical characteristics of voice and speech post-HNCa treatment may approximate normal, the perceptual end product is often non-normal.

Listener judgments may carry the greatest weight relative to “success” of treatment procedure.

Perceptual assessment captures multidimensional character of voice.
Some Interactions
EAI vs. DME Overall Severity ratings for TE speakers

$\text{DME Scale}$ vs. \text{Interval Scale}

$y = 4.752 \ln(x) - 15.478$

$R^2 = 0.7151$
EAI vs. DME Naturalness Ratings for TE Speakers

\[ y = 0.0319x + 1.2691 \]
\[ R^2 = 0.8533 \]
Figure 1 – EAI vs. DME pleasantness ratings for TE Speakers

\[ y = -9 \times 10^{-5}x^2 + 0.0501x + 1.3888 \]

\[ R^2 = 0.9202 \]
Figure 2 – EAI vs. DME acceptability ratings for TE speakers

\[ y = 0.0297x + 1.7576 \]

\[ R^2 = 0.8731 \]
Metathetic and Prothetic Continua

From Eadie & Doyle (2002b)
Laryngeal Cancer

Health Condition

Speech, shoulder function, smell, crying, laughing

Body Functions

Eating, bathing

Individual

Communication, social events

Societal

Activity

Participation

Contextual Factors

Environmental / Personal

Stigma, attitudes of family, friends; noise

Age, sex, ethnicity, education, SES, coping

Adapted from the ICF (WHO, 2001; Eadie, 2003)
Severity and Acceptability – Gender Issues
What is not known?

- The perceptual “character” of each method of alaryngeal speech (ES, TE, & EL)
- SI and VQ variables in other head & neck cancers
- A composite weighting of each
- The influence of other features
- The ultimate “capacity” for each method (i.e., “superior” speakers)

From: Doyle & Eadie, 2000a,b
“Speech [is] abnormal when it deviates so far from the speech of other people that it calls attention to itself, interferes with communication, or causes the speaker or his listeners to be distressed”

Charles Van Riper, 1978, p. 43

Social Impact and Disability
Thank you!

Questions?