

Developing Auditory-Perceptual Judgment Reliability in Otolaryngology Residents



Stephanie Misono¹

Albert L Merati¹

Tanya Eadie²

University of Washington, Seattle

¹Department of Otolaryngology/Head and Neck Surgery

²Department of Speech and Hearing Sciences



Introduction

- Auditory-perceptual methods are a vital component of voice evaluation
- Subject to error and variability¹
- Various methods (e.g., listener training with feedback; anchor samples) have been used to improve reliability of voice quality ratings^{2,3}

¹Kreiman et al., 1993; ²Eadie & Baylor, 2006; ³Chan & Yiu, 2006

Rationale for Study

- Effect of educational background and/or experience with dysphonia on reliability of judgments remains unclear
- Unknown reliability of judgments of dysphonia by otolaryngology residents
- Impact of listener training in residents also unknown

Purpose

To determine the effect of typical otolaryngology residency training and background on judgments of dysphonia

Experimental Questions

1. Are there differences in the reliability of judgments of dysphonia when judgments are made by otolaryngology residents and naive listeners?

Experimental Questions

1. Are there differences in the reliability of judgments of dysphonia when judgments are made by otolaryngology residents and naive listeners?
2. What is the effect of a brief training module on the reliability of resident listeners' judgments of dysphonia?

Methods: Part I

Voice samples:

- 24 recordings from individuals with a variety of laryngeal-based voice disorders
- Continuum of severity for dimensions of breathiness and roughness

Source: Voice Disorders Database (Kay Elemetrics, 1994)

Participants:

- 15 otolaryngology residents at UWMC
- 15 naive listeners

Participant Demographics: Part I

	Residents	Naive Listeners
Gender	11 M, 4 F	6 M, 9 F
Mean age (SD)	30.7 yrs (2.1 yrs)	25.3 yrs (8.8 yrs)
Laryngology exp.	1–54 months	None
Total	n=15	n=15

Data Collection: Part I

- Listeners familiarized with rating scales; definitions of roughness and breathiness (CAPE-V; ASHA, 2002)

Data Collection: Part I

- Listeners familiarized with rating scales; definitions of roughness and breathiness (CAPE-V; ASHA, 2002)
- Listened to speech sample through headphones; judged for roughness and breathiness on 10 cm visual analog scale



Data Collection: Part I

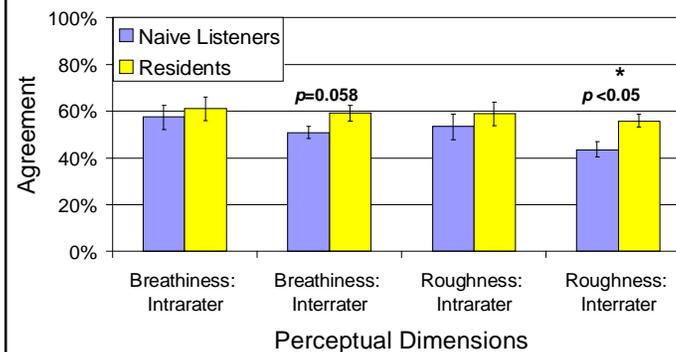
- Listeners familiarized with rating scales; definitions of roughness and breathiness (CAPE-V; ASHA, 2002)
- Listened to speech sample through headphones; judged for roughness and breathiness on 10 cm visual analog scale
- 6 stimuli repeated per dimension to calculate intrarater reliability



Data Analysis: Part I

- Dependent variables: Intrarater and interrater agreement
- “Agreement”: +/- 1 cm on 10 cm VAS
- 2 independent *t*-tests for each dimension to determine differences between resident and naive listener groups

Results: Part I



Results: Part I

- Weak correlation between years of otolaryngology training and reliability of listener judgments (largest $r = 0.29$)
- Educational background alone does not differentiate among residents
- Will brief training module improve judgments in this group?

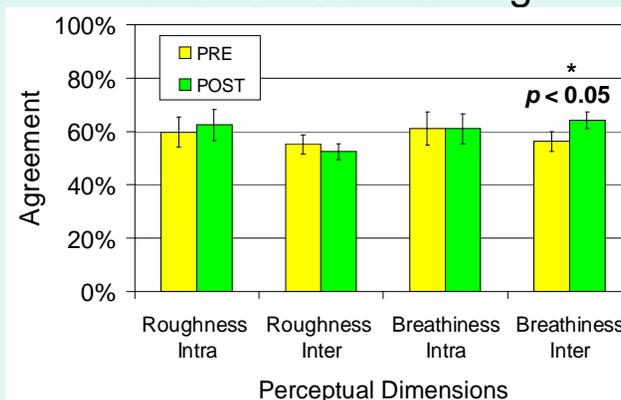
Methods: Part II

- Training module
 - 20 training items:
 - 18 dysphonic, 2 normal voices
 - Residents rated speech samples for roughness and breathiness
 - Immediate expert feedback on same stimuli (averaged from 3 SLPs; 11 yrs avg voice experience)
- Post-training test

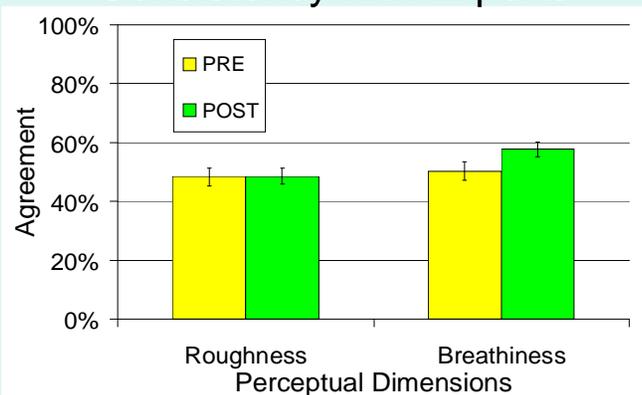
Data Analysis: Part II

- Dependent measures:
 - Intrarater agreement
 - Interrater agreement
 - Consistency with experienced listeners
- 3 matched pair *t*-tests per dimension to determine differences pre- to post-training

Results: Part II Pre vs Post-Training



Results: Part II Consistency with Experts



Conclusions: Part I

- Otolaryngology residents had better interrater agreement for judgments of breathiness and roughness than naive listeners

Conclusions: Part I

- Otolaryngology residents had better interrater agreement for judgments of breathiness and roughness than naive listeners
- Weak relationship between educational level (in residency program) and reliability

Conclusions: Part I

- Otolaryngology residents had better interrater agreement for judgments of breathiness and roughness than naive listeners
- Weak relationship between educational level (in residency program) and reliability
- Exposure to dysphonia or education in laryngology appears to help reliability of judgments, although not enough to differentiate residents as a group

Conclusions: Part II

- Improvement was greater for rating of breathiness, but not roughness

Conclusions: Part II

- Improvement was greater for rating of breathiness, but not roughness
- Breathiness has stronger acoustic correlates (Hillenbrand & Houde, 1996)

Conclusions: Part II

- Improvement was greater for rating of breathiness, but not roughness
- Breathiness has stronger acoustic correlates (Hillenbrand & Houde, 1996)
- Limitations:
 - No control group
 - Short time for training

Conclusions: Part II

- Improvement was greater for rating of breathiness, but not roughness
- Breathiness has stronger acoustic correlates (Hillenbrand & Houde, 1996)
- Limitations:
 - No control group
 - Short time for training
- Future implications: Standardized training in evaluation of dysphonia

Acknowledgments

- Tanya Eadie, Ph.D. CCC-SLP
- Albert L. Merati, MD FACS
- Amanda Politziner, BA
- Derek Wright, BS
- Participants

Questions

