Toward an Objective Multidimensional Evaluation of Voice Quality in Head And Neck Cancer

Xinyu Zhang

Under the supervision of

Dr. Rob van Son

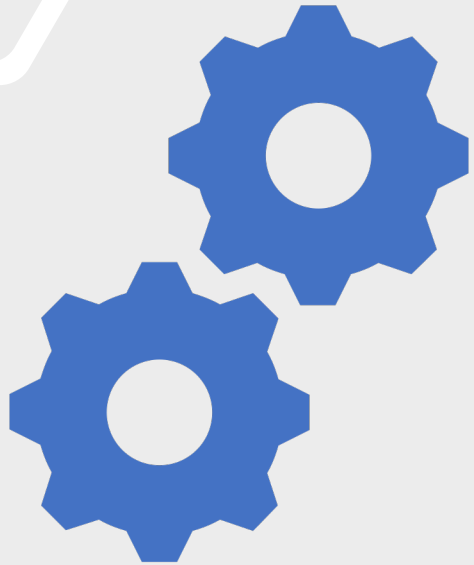
University of Amsterdam

Amsterdam Center for Language and Communication

The Netherlands Cancer Institute

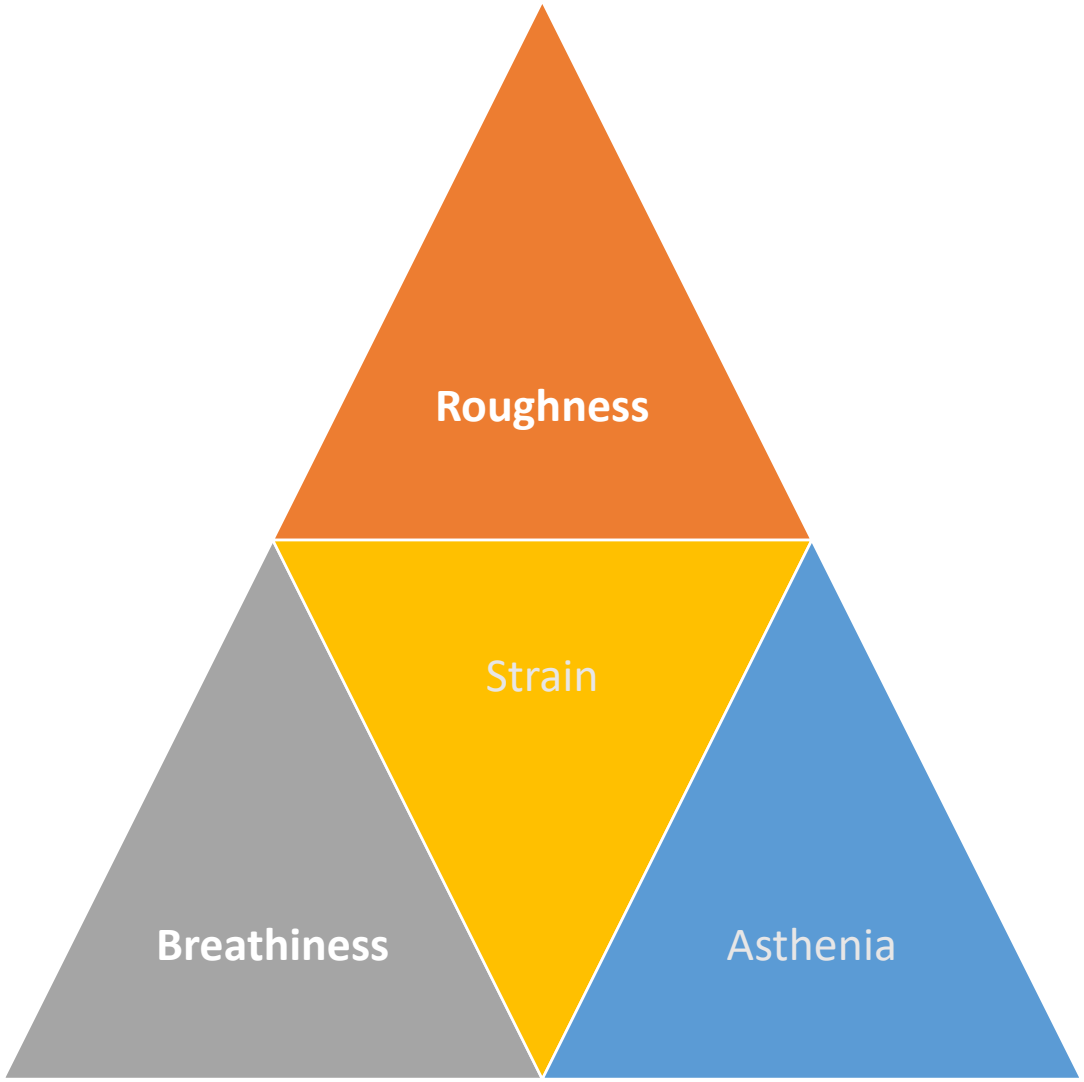
Topic and Goals

- Acoustic metrics for the purpose of voice quality judgement
- For different aspects of voice quality
- That includes continuous speech
- Target population: Individuals undergone treatment for laryngeal carcinoma



Voice Quality

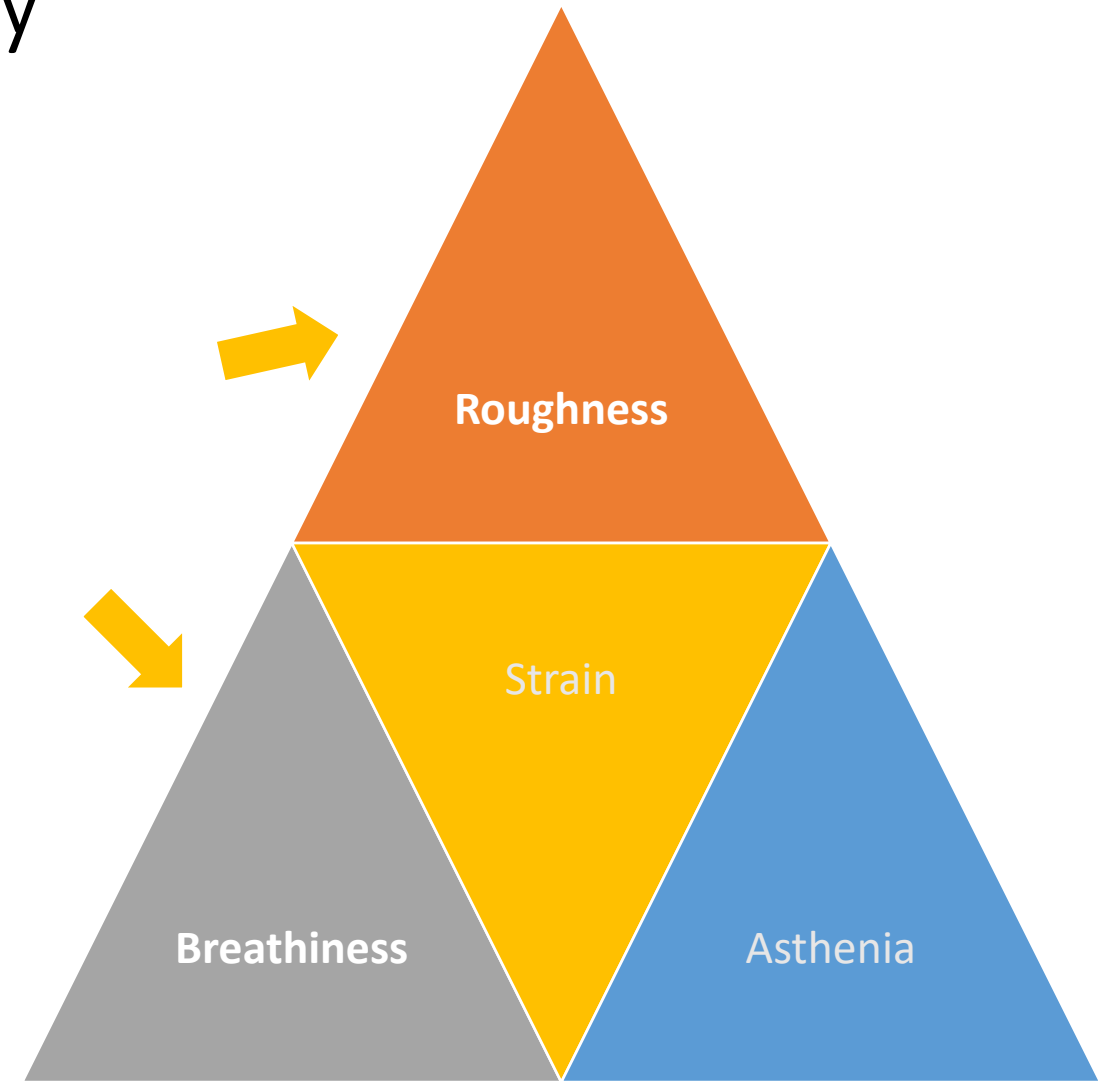
- Multi-dimensional
- A perceptual phenomenon
- Not be easily measurable



Change of Voice Quality in the Treatment of Laryngeal Carcinoma

- Dryness
- Loss of muscle bulk
- Fibrosis
- Keratosis
- Reinke's Edema

- Irregular vibration
- Lack of proper closure



Evaluation of Voice Quality

Subjective

Auditory-perceptual judgment by speech pathologist (Nawka et al., 1994; Hirano, 1981; Kempster et al., 2009, etc.)

Objective

-Direct observation of vocal fold vibration (e.g. Deliyski et al., 2008)

-Aerodynamic measurements (Lim et al., 2016)

-Multi-parametric acoustic measurements (e.g. Maryn et al., 2010)

Limitations of Current Acoustic Evaluation Processes

Perturbation measures

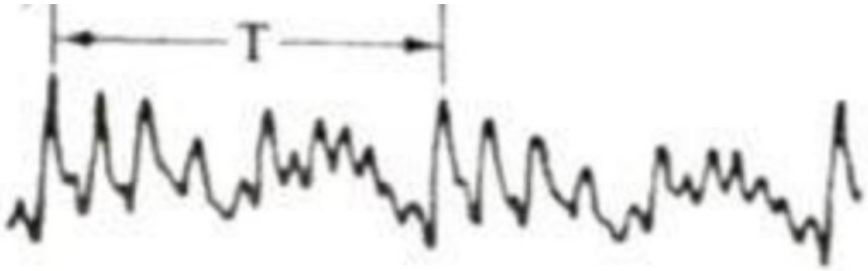
Only work well with sustained vowels

Require selection and segmentation by hand

Often only give a score on the general voice quality

Problems with Sustained Vowels

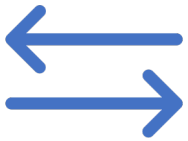
- Getting them
- Not usually what is judged and perceived in natural speech
- Do not reflect vocal characteristics of the speaker
- Do not show irregularities in e.g. rapid onset/termination
- Requires selecting a stable part of the vowel
  - => Severely dysphonic voices can’t be examined
  - => causing a gap



Breathiness vs. Roughness

- Both having irregular periodicity
  - Both high in jitter and shimmer
  - Both low in HNR
- They co-occur and correlate!

Methods



Listening Experiment

- Professionally trained listeners (practicing SLPs)
- Continuous speech (~5 sec / stimuli)
- Ratings on roughness and breathiness

Acoustic Measurements

- Continuous speech and sustained vowels
- 100 recordings
- Perturbation measures on vowels
- Spectral- and Cepstral- based measurements

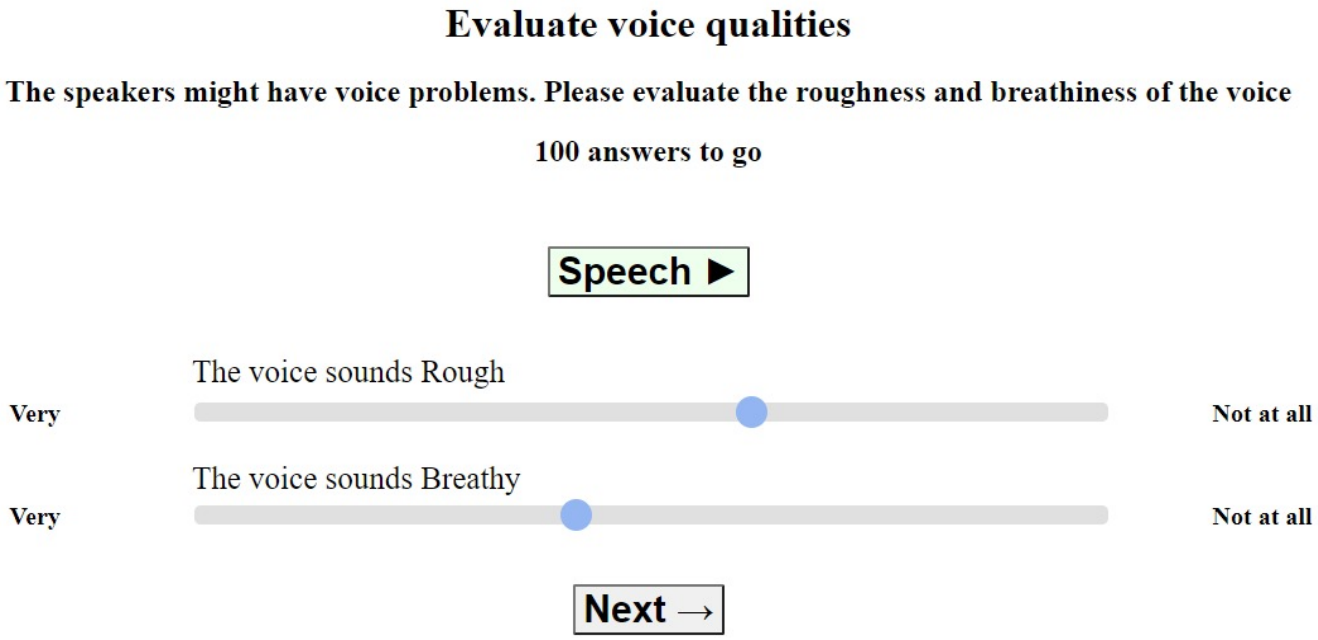
The Listening Experiment

Stimuli

- 45 patients of laryngeal carcinoma (stage CIS-T2N0M0)
- Reading a content-neutral text of ~ 150 words, during usual care
- Recordings at pre-treatment, 6 months post-treatment, and 12 months post-treatment
- 100 recordings in total

The Listening Experiment

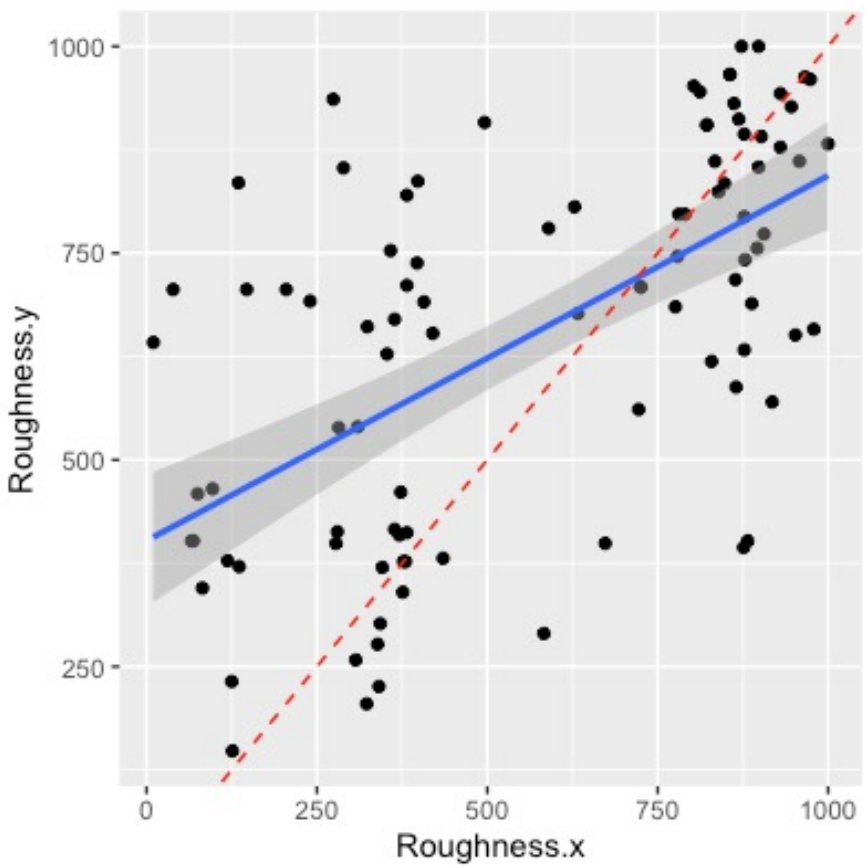
Participants & Procedures

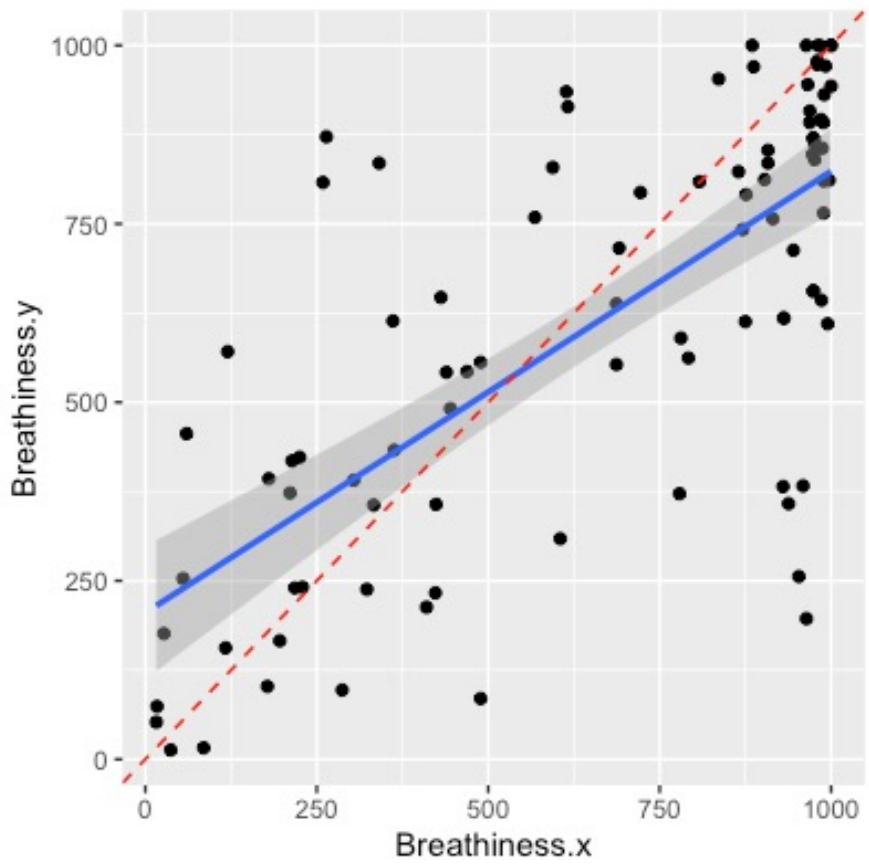


- 2 SLPs from the NKI
- 5 seconds of read speech per stimulus
- Randomized
- “Breathiness” and “Roughness” as defined in the GRBAS (Hirano, 1981)
- Visual Analog Scale (0-1000)
- 100 stimuli (incl. 4 practice items)

The Listening Experiment

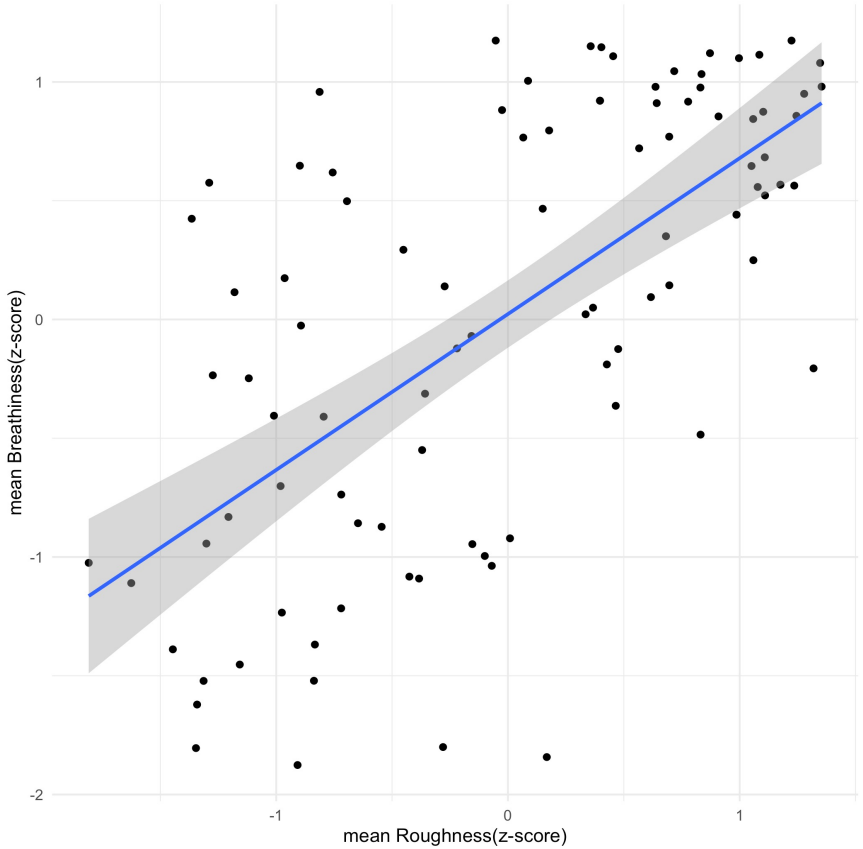
Between-Evaluator Agreement



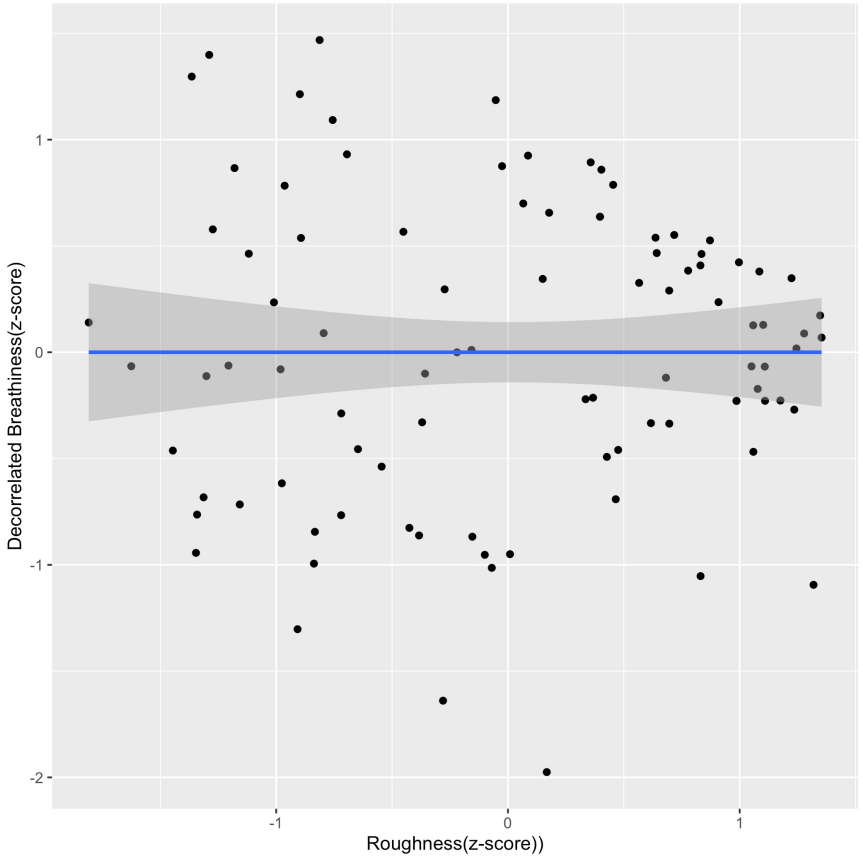


The Listening Experiment

Confounding and Decorrelation



Before



After

Acoustic Measurements

Recordings

- Same 45 patients from the listening experiment
- Read-speech at different stages of treatment (4 sec x 100 recordings in total) -> extracted voiced segments only
- A sustained corner vowel (3 sec x 100 recordings in total)

Subject ID	Time point	Files
S1	T1	continuous speech (4 sec) + sustained vowel (3 sec)
	T2	continuous speech (4 sec) + sustained vowel (3 sec)
	T3	continuous speech (4 sec) + sustained vowel (3 sec)
S2	T1	continuous speech (4 sec) + sustained vowel (3 sec)
	T2	continuous speech (4 sec) + sustained vowel (3 sec)
	T3	continuous speech (4 sec) + sustained vowel (3 sec)
⋮		⋮
S45	T1	continuous speech (4 sec) + sustained vowel (3 sec)
	T2	continuous speech (4 sec) + sustained vowel (3 sec)
	T3	continuous speech (4 sec) + sustained vowel (3 sec)

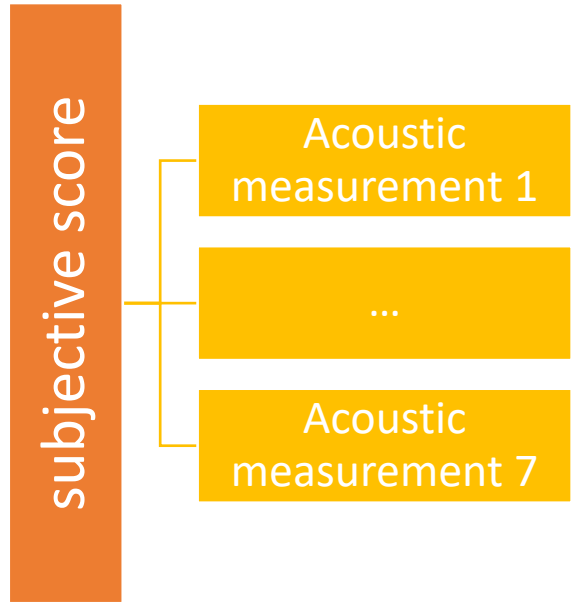
Acoustic Measurements

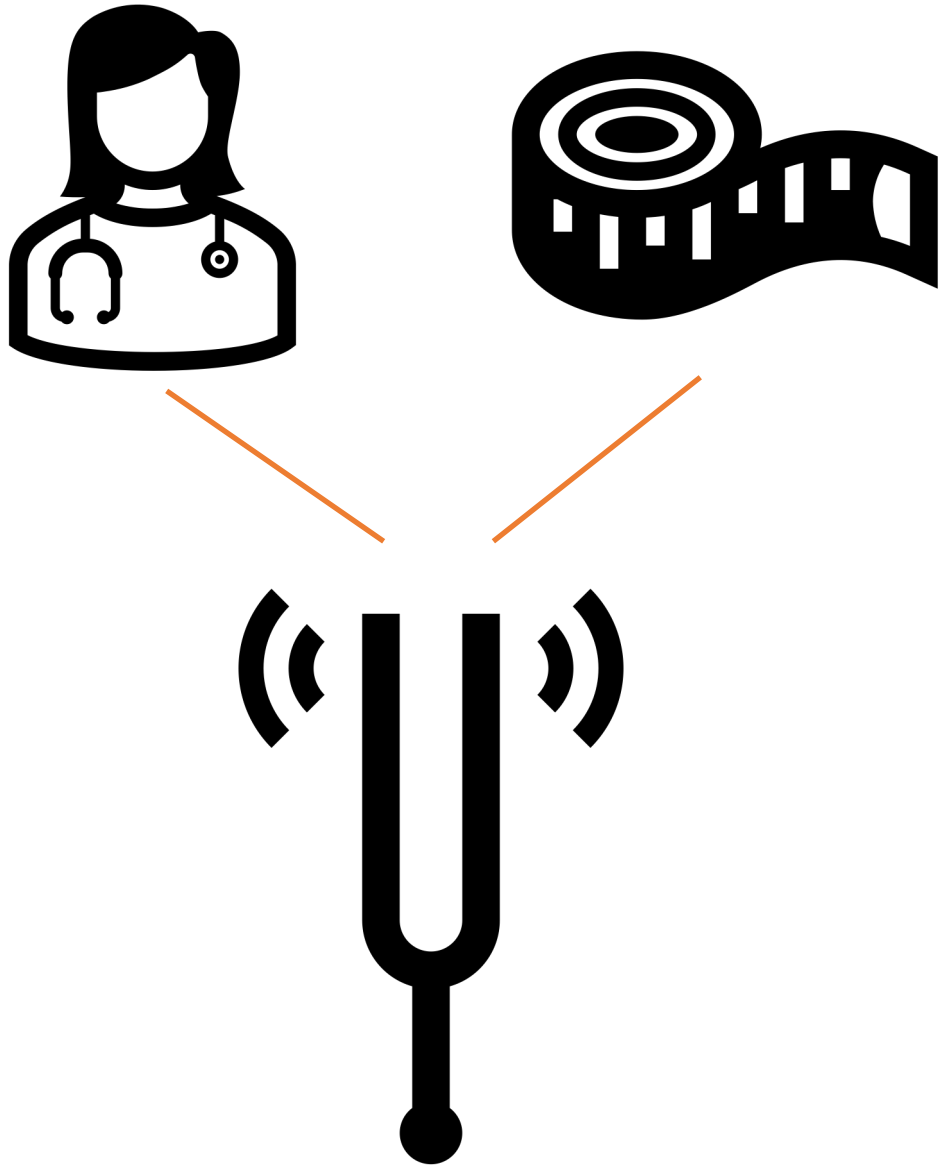
Metrics

Measurement	Description	Type
CPPS	Smoothed Cepstral Peak Prominence	Quefrency-based
HNR	Harmonics-to-Noise Ratio	Glottal noise measure
Jitter	Cycle-to-cycle frequency variation	Time-based
Shimmer	Cycle-to-cycle amplitude variation	Time-based
ShdB	Shimmer in dB	Time-based
Slope	Slope of the Long-Term Average Spectrum (LTAS)	Frequency-based
Tilt	Tilt of the trend line through the LTAS slope	Frequency-based

Statistical Analysis

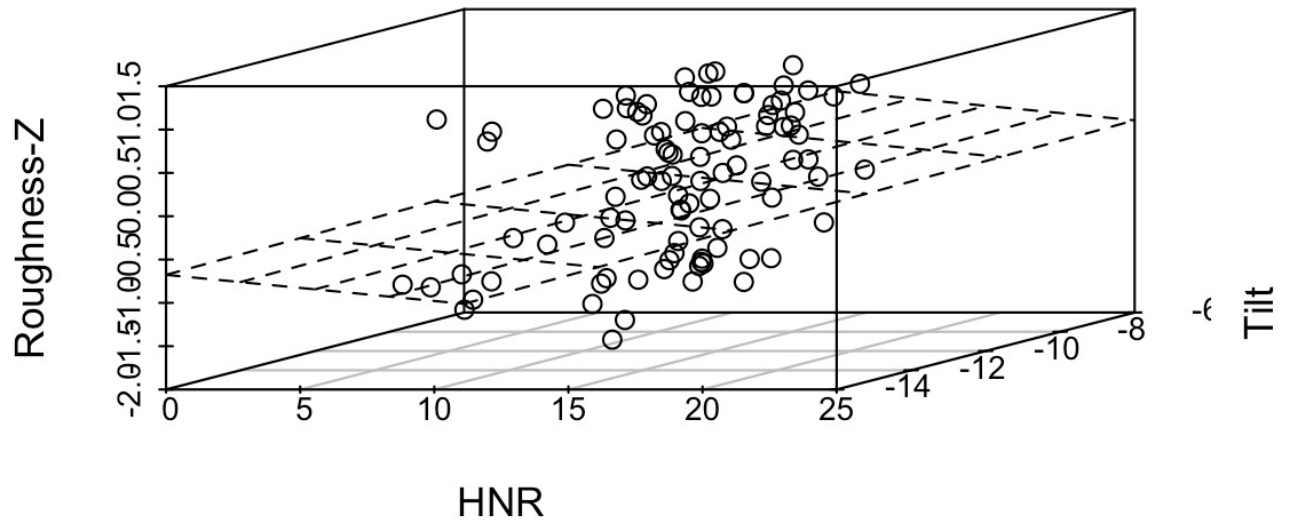
- Perceptual scores as ground truth
- Step-wise linear regression





Results

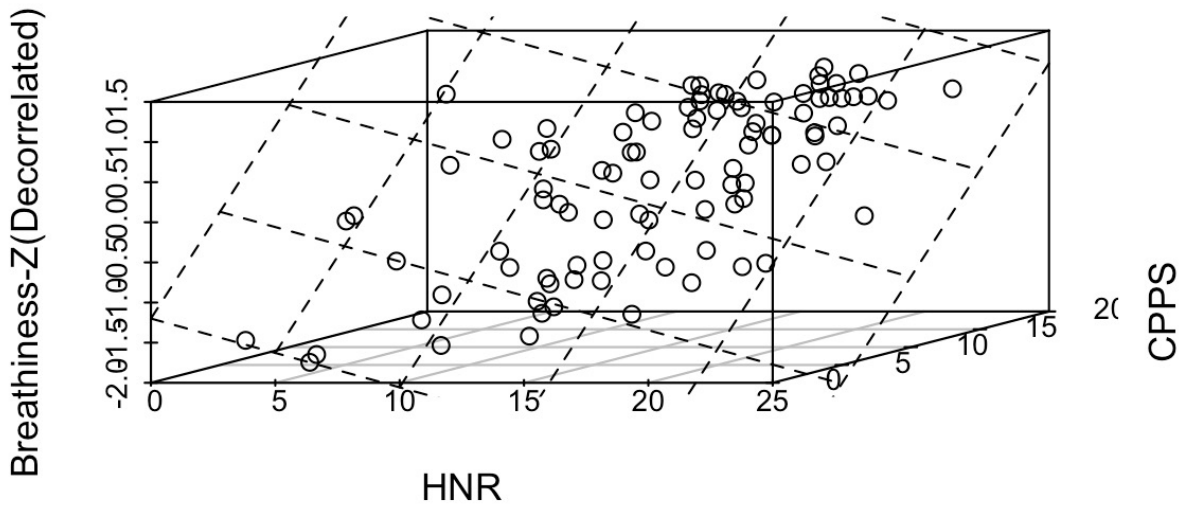
Roughness



$$\widehat{meanRoughnessZScore} = -2.81821 + 0.08474 * HNR - 0.15297 * Tilt$$

Results

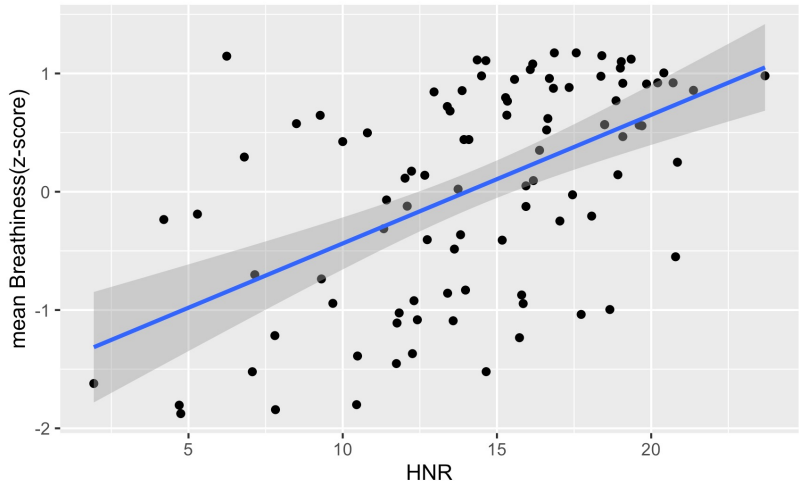
Decorrelated Breathiness

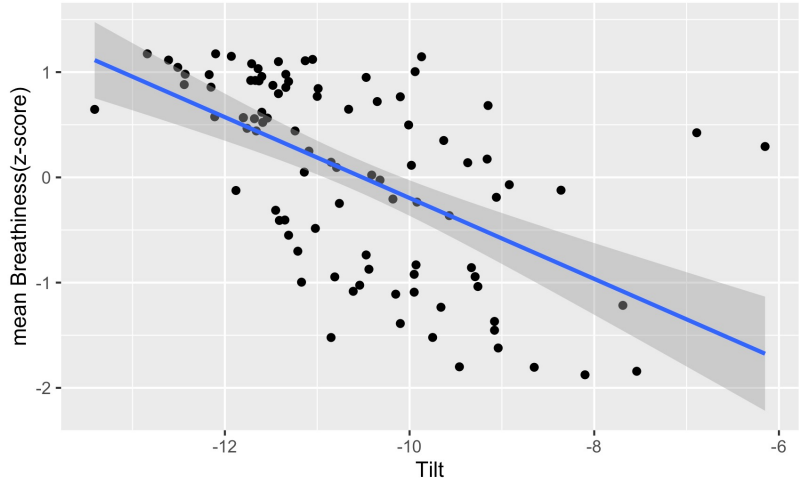


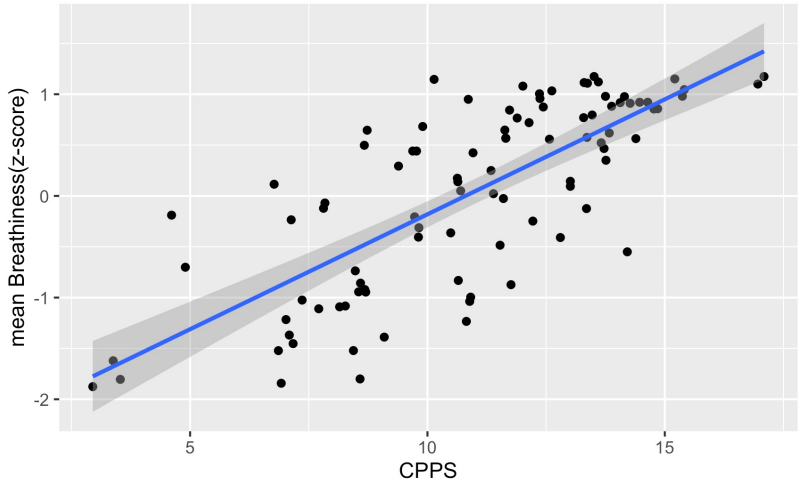
$$\widehat{meanDecorr.BreathinessZScore} = -1.19941 - 0.08569 * HNR + 0.22166 * CPPS$$

Results

Reconstructed Breathiness



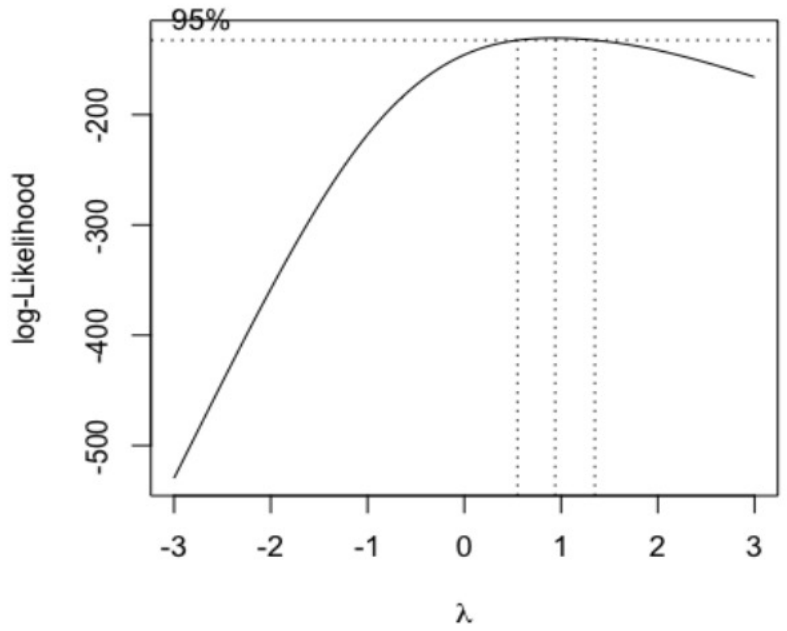




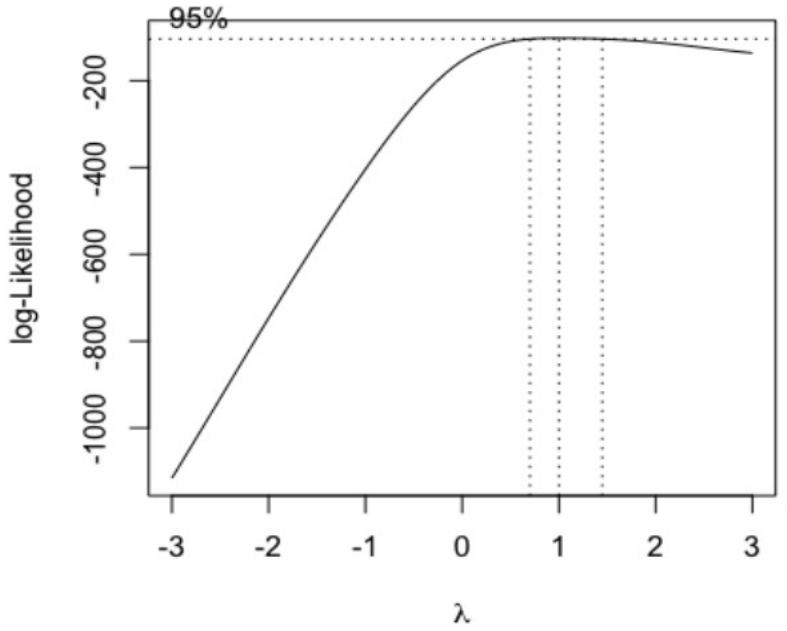
$$\widehat{meanBreathinessZScore} = -3.18284 - 0.04923 * HNR + 0.25450 * CPPS - 0.10715 * Tilt$$

Results

Linearity & Model Assumptions



(a) Roughness



(b) Decorrelated-Breathiness

The best λ and their 95% CI for the two models

The Explainable and the Explained

Roughness:

- *Adjusted* R^2 = 0.364,
- 4% variance caused by between rater disagreement

Explained:

$$\frac{0.364}{1-0.04} = 37.92\%$$

1/2 variables measured on continuous speech

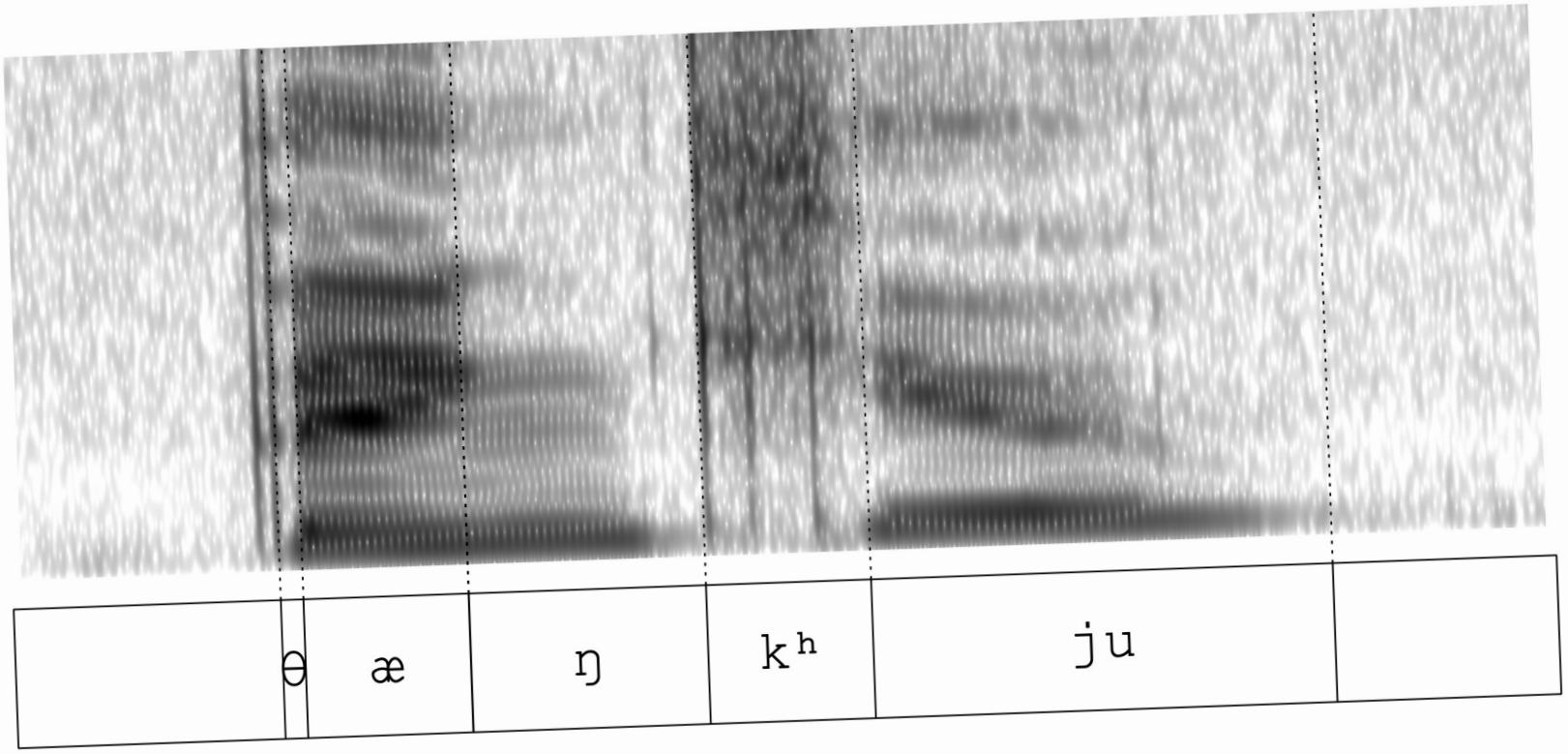
Breathiness:

- *Adjusted* R^2 = 0.583
- 2.3% variance caused by between rater disagreement

Explained:

$$\frac{0.583}{1-0.023} = 59.67\%$$

2/3 variables measured on continuous speech


θ
æ
ŋ
kʰ
ju

References

- Deliyski, D. D., Petrushev, P. P., Bonilha, H. S., Gerlach, T. T., Martin-Harris, B., & Hillman, R. E. (2008). Clinical implementation of laryngeal high-speed videoendoscopy: challenges and evolution. *Folia Phoniatrica et Logopaedica*, *60*(1), 33-44.
- Hirano, M. (1981). Psycho-acoustic evaluation of voice. *Clinical examination of voice*, 81-84.
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*.
- Lim, J. Y., Choi, J. N., Kim, K. M., & Choi, H. S. (2006). Voice analysis of patients with diverse types of Reinke's edema and clinical use of electroglottographic measurements. *Acta oto-laryngologica*, *126*(1), 62-69.
- Maryn, Y., De Bodt, M., & Roy, N. (2010). The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *Journal of communication disorders*, *43*(3), 161-174.
- Nawka, T., Anders, L. C., Wendler, J., & Anders, L. (1994). Die auditive Beurteilung heiserer Stimmen nach dem RBH-System.