Perceptual and Acoustic Assessment of Strained Voice Quality Using Synthetically Modified Voice Samples

Park, Y., Díaz Cádiz , M. E., Nagle, K. F., & Stepp, C. E.

BOSTON UNIVERSITY

STEPP LAB
Sensorimotor Rehabilitation Engineering

Introduction

Strained Voice Quality

- Voice quality: “attributes of a sound other than its pitch and loudness that allow a listener to judge that two sounds are the same or different” (ANSI, 1960)
- **Strain**: “perception of **excessive vocal effort** (hyperfunction)” (Kempster et al., 2009)
- One of the common features of hyperfunctional voice disorders (e.g., nodules, muscle tension dysphonia), laryngeal dystonia, etc.
- Mainly assessed with subjective, auditory-perceptual evaluation (Oates, 2009)
- **Lack of knowledge in acoustic features of strain**

Acoustic Features of Strain

- Increased perceived strain has been associated with:
  1) Increased spectral energies at higher harmonic frequencies (Anand et al., 2019; Bergan et al., 2004)
  2) Decreased **relative fundamental frequency** (**RFF**; more information on Appendix)
     - RFF quantifies the short-term variations of fundamental frequency (f_o) in voiced-voiceless consonant-voiced productions (e.g. /ifi/)
     - RFF was observed to be decreased in individuals with hyperfunctional voice disorders (Stepp et al., 2010, 2011)
  3) Increased **mid-to-high frequency noise** (Hirano, 1981; Klatt & Klatt, 1990; Lowell et al., 2012)
- **Not clear whether RFF and mid-to-high frequency noise directly contribute to strain**

Purpose and Hypotheses

To understand the effect of **RFF** and **mid-to-high frequency noise** on strain by **synthetically modifying** these features

Hypothesis 1: **Lowering RFF** in voice samples would **increase strain**

Hypothesis 2: **Raising RFF** in voice samples would **decrease strain**

Hypothesis 3: **Adding mid-to-high frequency noise** to voice samples would **increase strain**

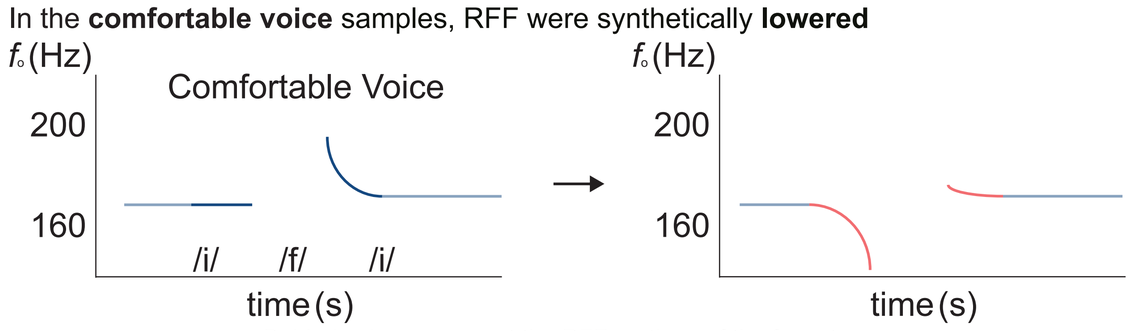
Methods

Original Voice Samples

- /ifi/ recordings of eight speakers (4 females; mean age = 32.6 years) with typical voices
- Both **comfortable voice** and **maximum vocal effort samples** for each speaker (8 speakers x 2 vocal effort levels = 16 original samples)

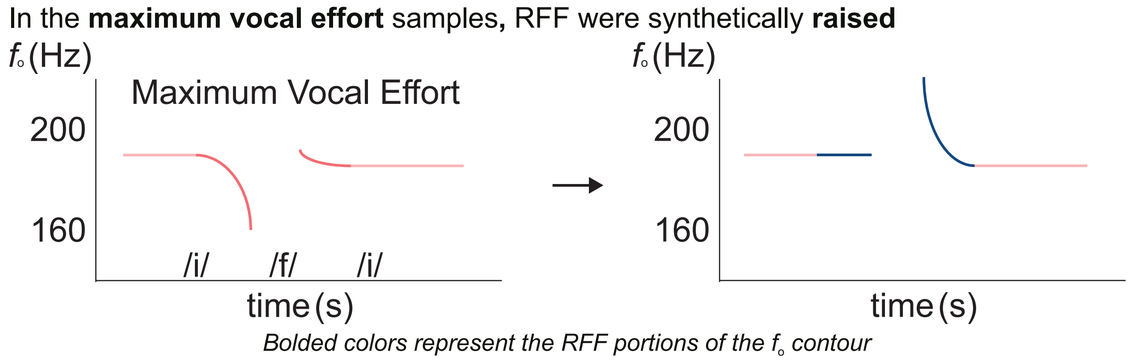
RFF Modification

- In the **comfortable voice** samples, RFF were synthetically **lowered**



Bolded colors represent the RFF portions of the f_o contour

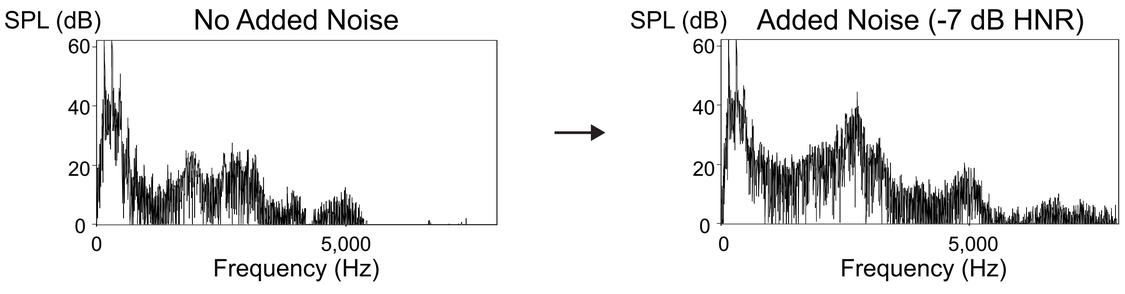
- In the **maximum vocal effort** samples**,** RFF were synthetically **raised**



Bolded colors represent the RFF portions of the f_o contour

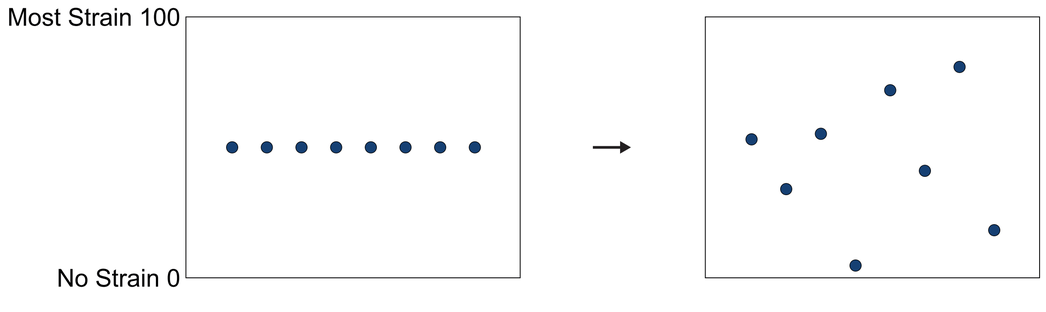
Mid-to-High Frequency Noise

Added to the sample to **decrease harmonics-to-noise ratio (HNR) by 7 dB**



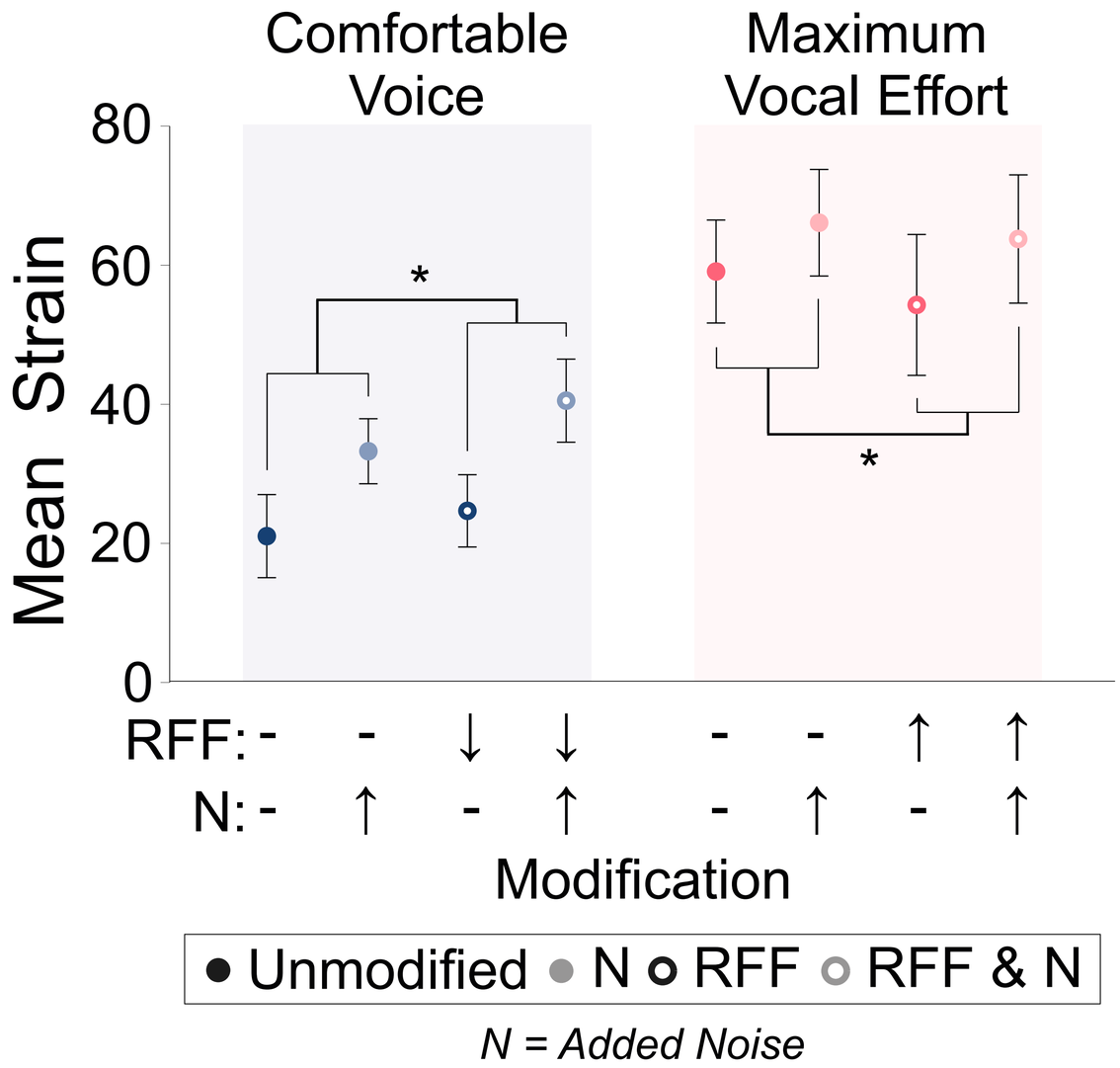
Visual Sort-and-Rate Task

- 20 listeners (10 females; mean age = 22.0 years)
- 64 stimuli = 16 unmodified + 16 RFF-modified + 32 added noise



Results

- The effect of **the interaction between RFF and vocal effort levels** (p = 0.003, η_p^2 = 0.74)
  - Lowering RFF increased strain (p < 0.001)
  - Raising RFF decreased strain (p = 0.003)
- The effect of **mid-to-high frequency noise** (p < 0.001, η_p^2 = 0.97)



N = Added Noise

Discussion

- Directly supports **RFF** and **mid-to-high frequency noise** as an **acoustic contributors to strain** using synthetically modified samples
- **Multiparametric tools** incorporating multiple acoustic features of strain

References

- Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2019). Objective Indices of Perceived Vocal Strain. *Journal of Voice*, *33*(6), 838-845.
- ANSI (1960) S1.1-1960, Acoustical terminology, New York: American National Standards Institute.
- Bergan, C. C., Titze, I. R., & Story, B. (2004). The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice. *Journal of Voice*, *18*(3), 305-317.
- Hirano, M. (1981). Psycho-acoustic evaluation of voice: GRBAS scale for evaluating the hoarse voice. In *Clinicain Examination of Voice*. New York: Springer-Verlag.
- Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., 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*, *18*(2), 124-132.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of Acoustical Society of America, 87*(2), 820-857.
- Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *The Annals of Otology, Rhinology, and Laryngology, 121*(8), 539-548.
- Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatrica et Logopaedica*, 61(1), 49-56.
- Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53*(5), 1220-1226.
- Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E. (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 54*(5), 1260-1266.

Acknoledgement

This work was supported by grant DC015570 from the National Institute on Deafness and Other Communication Disorders and a Dudley Allen Sargent Research Fund Grant from Boston University.

NIDCD
National Institute on Deafness and Other Communication Disorders

Appendix

Relative Fundamental Frequency

Instantaneous f_os of the 10 voicing cycles before and after a voiceless consonant are normalized by the f_os of the cycles furtherst from the consonant (f_{ref}; offset cycle 1 and onset cycle 10) with the equation: RFF (ST) = 12 × $\log_2(f_o/f_{ref})$