

Toward a Psychoacoustic Model of Spectral Noise in the Voice Source

Rosario Signorello⁽¹⁾, Nari Rhee, Bruce R. Gerratt⁽¹⁾, and Jody Kreiman⁽¹⁾

⁽¹⁾Bureau of Glottal Affairs, Department of Head and Neck Surgery,
University of California,
jkreiman@ucla.edu

Introduction

Previous studies (Kreiman *et al.*, 2015) have proposed a psychoacoustic model of voice quality that includes a four-parameter spectral-domain model of the harmonic voice source (H1-H2, H2-H4, H4-the harmonic nearest 2 kHz, and that harmonic to the harmonic nearest 5 kHz). In this paper we report a parallel psychoacoustic study of the spectral shape of the inharmonic source. Our goal is two-fold: to provide a framework for nuanced measurement of vocal noise, and to improve understanding of the interaction between harmonic and inharmonic aspects of the voice source in determining voice quality.

Methods: Acoustic Analysis

120 steady-state /a/ vowels (60 female speakers; 15 each normal, mildly, moderately, and severely pathologic) were copy-synthesized using analysis-by-synthesis. The noise spectrum was extracted from each synthesized vowel and downsampled to 111 points. These data were analyzed using principal component analysis (PCA) with varimax rotation to derive a set of factors describing noise spectral shapes across voices.

Results

For both female and male voices, PCA revealed four factors that accounted for most of the variance in the underlying data (females: 87%; males: 91%). For female speakers, these factors divided the spectrum into components ranging from 0-1400 Hz, 1400-2800 Hz, 2800-4200 Hz, and 4200-5000 Hz. For male speakers, the spectrum was divided into components ranging from 0-600 Hz, 600-2200 Hz, 2200-3800 Hz, and 3800-5000 Hz.

Perceptual Evaluation

Two synthetic versions of each voice were created: one with the original noise spectrum, and one with the noise spectrum smoothed with a 4-piece model corresponding to the ranges described above. To determine if the four-piece model adequately captures the perceptually-important features of the inharmonic source, listeners compared these two synthetic tokens to each other and to the original natural voice sample. If the spectral model is adequate, then

the two synthetic tokens should be indistinguishable from one another and from the original voice sample. Comparison to the original voice sample establishes the adequacy of the synthesis on which the analysis was based; comparison of the two synthetic tokens establishes the adequacy of the 4-parameter model. This experiment is currently underway.

Conclusions

Preliminary results suggest that the psychoacoustic model is sufficient to characterize perceptually-important aspects of the spectral shape of the inharmonic voice source for normal and pathologic steady-state phonation.

Acknowledgements

This study was supported by NIH grant DC01797 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, United States.

References

Kreiman, J., Gerratt, B. R., Signorello, R., and Rastifar, S. (2015). "Adequacy of a four-parameter spectral model of the voice source". In *Journal of the Acoustical Society of America*, 137(4): 2266.