Perceptual and acoustic assessment of strain using synthetically modified voice samples

Yeonggwang Park[1], Manuel Diaz-Cadiz[1], Kathleen F. Nagle[2], and Cara E. Stepp[1,3,4]

[1]Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA

[2]Department of Speech-Language Pathology, Seton Hall University, South Orange, NJ

[3]Department of Biomedical Engineering, Boston University, Boston, MA

[4]Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA

Purpose: Strain is a voice quality related to perceived excessive vocal effort frequently present in voice disorders. A primary goal of voice therapy for these voice disorders is reduction of strain, but its assessment has been difficult due to the weak reliability of auditory-perceptual evaluation. Previous studies have struggled to identify acoustic measures that reliably correlate with strain. However, these studies have used natural voice samples, which potentially contained variations in all of their acoustic features. To study the specific relationship between acoustic and perceptual measures, only the measures of interest should be modulated; other measures should be held constant. In this study, we evaluate the relationship between strain and two acoustic features, relative fundamental frequency and aspiration noise, by synthetically manipulating them in voice samples.

Method: From an existing database we obtained recordings of eight speakers who produced RFF stimuli with the instruction to produce them using either a comfortable voice or with maximum vocal effort. In the recordings produced with maximum vocal effort, strain was higher and RFF was lower. The STRAIGHT (Speech Transformation and Representation using Adaptive Interpolation of weighted spectrum) algorithm (Kawahara, 2006) was used to manipulate RFF in these recordings. RFF values from the samples produced with a comfortable voice were modified to be lower to examine if strain would increase as a result,

and RFF values from the maximum vocal effort samples were modified to be higher to examine if strain would decrease as a result. Aspiration noise was added to both RFF-modified and -unmodified samples via Praat's plug-in function, "breathiness." In this abstract we present the preliminary results from ten inexperienced listeners (an additional ten listeners will be recruited). Listeners rated the strain of the synthesized voice samples with a visual sort-and-rate (VSR) task. Once the final dataset of 20 listeners is complete, we will perform a three-way ANOVA on the mean strain of the samples with the factors, vocal effort level, RFF manipulation, aspiration noise, and their interactions.

Results: Based on 10 listeners, decreasing RFF of the comfortable voice samples resulted in slight increases in strain, whereas increasing RFF of the maximum vocal effort samples resulted in slight decreases in strain (Figure 1). The added aspiration noise had the effect of slightly increased strain ratings in all samples (Figure 1).

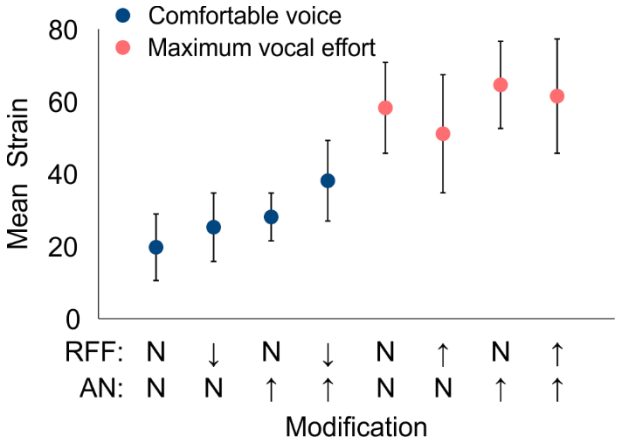

Figure 1 *Mean strain scores of the eight samples under a same modification category from the VSR task (abbreviations: AN= aspiration noise; N=no modification; ↓=decrease; ↑=increase). Error bars represent the standard deviations.*

Conclusion: Based on these initial results, RFF and aspiration noise may affect strain, although their effects may be small.

Acknowledgments: This work was supported by grant DC015570 from the National Institute on Deafness and Other Communication Disorders.

References

Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoust. Sci. & Tech., 27*(6), 349-353.