HIGH-SPEED VIDEOENDOSCOPY IN THE CLINICAL ASSESSMENT OF VOICE DISORDERS: THE EFFECT OF SPEED ON VIBRATORY FEATURES

Shannon R. Batson, Kimberly S. Hufnagel, Habib J. Moukalled, Dimitar D. Deliyski

University of South Carolina, Department of Communication Sciences and Disorders, Columbia, SC

ABSTRACT:

This research helps in establishing a clinical protocol for the assessment of voice disorders using high-speed videoendoscopy (HSV). It was hypothesized that reduced HSV frame rates affect the accuracy and reliability when visually evaluating the following vocal-fold vibratory features: mucosal wave, amplitude and phase asymmetry, aperiodicity, glottal edge roughness, vocal-fold contact, and mucus bridges.. One habitual- and one high-pitch sample were taken from two male and two female subjects selected from the dataset that best represented each vibratory feature. The selected 16,000 fps samples were then down-sampled fully emulating the characteristics of the camera at sixteen lower-rate denominations. Three raters performed paired comparisons of 816 HSV recordings sampled at various frame rates. This poster reports the findings about the frame rates at which degradation in each feature was first noted and where the feature could no longer be judged. Research supported by NIH/NIDCD grant R01DC007640.

PURPOSE:

The long-term purpose of this research is to establish a clinical protocol and to evaluate the reliability, validity, and relevance of high-speed videoendoscopy (HSV) in the clinical assessment of voice disorders. For this project, we primarily addressed how HSV frame rate impacts visual evaluations of vocal fold vibratory features. It was hypothesized that lowering the HSV frame rate directly affects the accuracy and reliability of visual assessment. This project sought to identify the frame rates at which clinical features are affected by the speed of videoendoscopy and to identify the adequate frame rates for using HSV in a clinical setting.

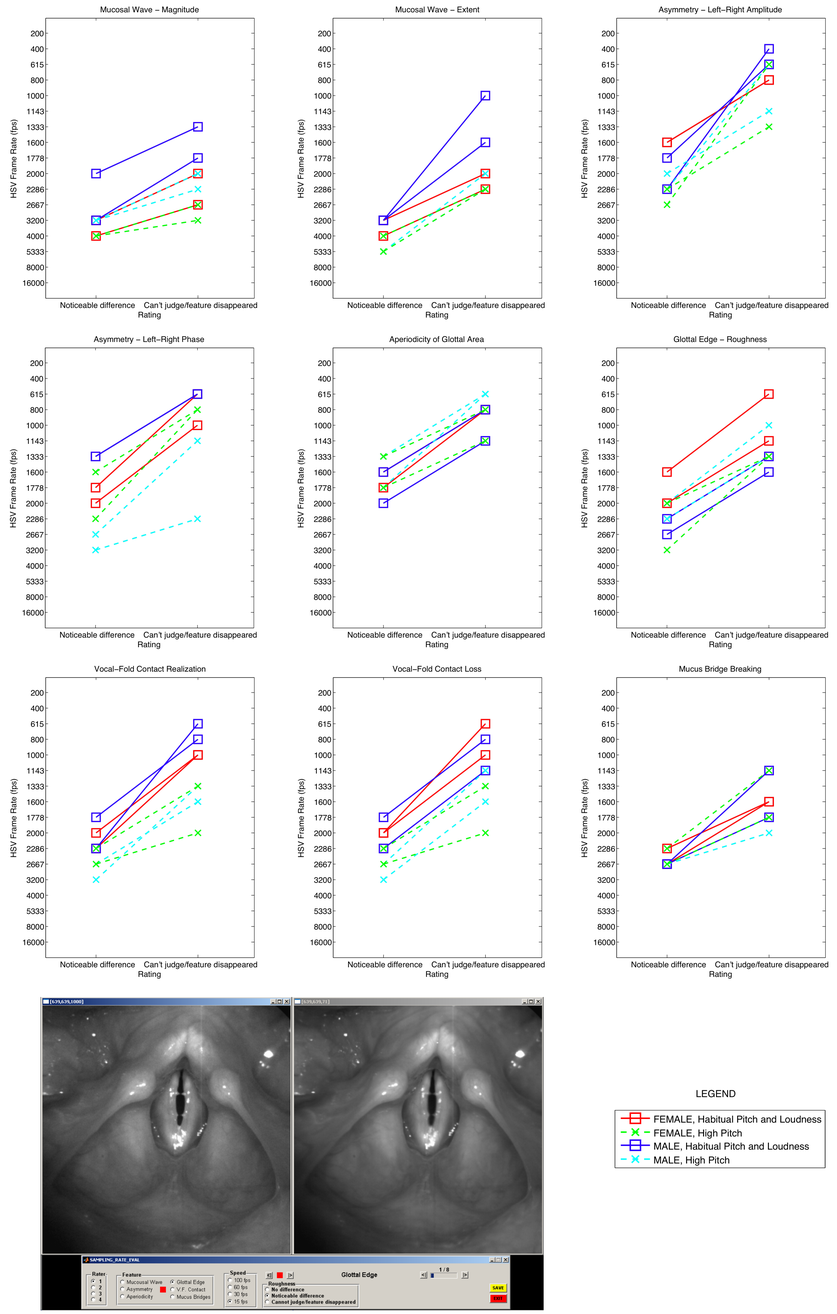
PARTICIPANTS:

Fourteen patients and fourteen vocally-normal speakers were recorded using HSV at 16,000 frames per second (fps) as they produced the vowel /i/ sustained at habitual pitch and loudness, and at higher pitch. One habitual- and one high-pitch sample were then taken from two male and two female subjects selected from the dataset that best represented each vibratory feature.

Three raters, all students of the University of South Carolina, judged 9 vocal-fold vibratory features. Two of the raters were graduate Speech-Language Pathology students with experience in the area of voice. The third rater was a Computer Science doctoral student with knowledge in biomedical engineering, computer vision, and theory of computation. The raters were trained on the nature of vibratory features within the vocal folds, and how to evaluate these features.

PROCEDURE:

The selected 16,000-fps recordings were down-sampled to accurately emulate 17 different denominations of the frame rate descending from 16,000 to 200 fps. Custom software (see figure) allowed for parallel visualization of each fps denomination next to the 16,000-fps reference. The presentation of the samples was randomized so that the raters were unaware of the gender, pitch and frame rate they were judging. Prior to judging, the raters underwent training sessions to learn how to identify and to calibrate the sensitivity of judging each feature. The raters were trained by a researcher with an extensive background in voice. All rating sessions occurred within a week of training.



PROCEDURE CON'T:

The features evaluated were: magnitude and extent of mucosal wave, left-right amplitude and phase asymmetry, glottal area aperiodicity, glottal edge roughness, realization and loss of vocal-fold contact, and breaking of mucus bridges. Features were selected by considering their impact on the evaluation of voice disorders and sensitivity to change in frame rate. Using the software, each of the three raters made visual judgments for each feature selected for a total of 816 ratings each. The features were judged to detect the point at which: (1) a difference between the reference and the degraded sample was first noticed; and (2) the feature disappeared or became hard to visually rate.

RESULTS:

The ratings were analyzed by feature and sample. The results varied greatly across features. The raters varied in their sensitivity of judging. The figures show the median score across raters for each feature by gender and pitch. The findings revealed that for the features of mucosal wave, asymmetry, glottal edge roughness, and vocal-fold contact the changes were first noticed in the high pitch female samples. In general, habitual pitch and loudness samples were least sensitive to the down-sampling. In the case of aperiodicity, however, habitual pitch showed more sensitivity than high pitch. Both, realization and loss of vocal-fold contact were found to have similar ratings of sensitivity. Regarding the breaking of mucus bridges, a difference was noted in all samples at nearly the same frame rate, but male high pitch samples could no longer be judged at frame rates higher than other samples.

CONCLUSION:

High pitch phonation caused vocal-fold features to be generally more sensitive to a reduction in frame rate. Judgments of habitual pitch phonations were less dependent on the speed of recording. Based on these results, many of the features cannot be adequately judged using current commercial clinical HSV equipment, typically operating in the range of 2,000-4,000 fps. Clinicians need access to equipment that can provide adequate frame rates to assess the vibratory features of voice. Further investigation needs to be conducted in order to establish the inter- and intra-rater reliability. Additional rating from a highly skilled professional in the area of voice needs to be conducted in order to increase the reliability of this study and establish a consensus. Mucosal wave is the feature most demanding in terms of frame rate, probably requiring rates above 5,000 fps. Further investigation of the relationship between pitch and required frame rate will aid in refining current clinical protocols.

REFERENCES:

• Bonilha, H.S., Deliyski, D.D., & Gerlach, T.T. (2008). Phase asymmetries in normophonic speakers: Visual judgments and objective findings. *American Journal of Speech-Language Pathology* 17(4), 367-376.
• Bonilha, H.S. & Deliyski, D.D. (2008). Period and glottal width irregularities in vocally normal speakers. *Journal of Voice* 22(6), 699-708.
• Bonilha, H.S. & Deliyski, D.D. (2008). Vocal fold mucus aggregation in vocally normal speakers. *Logopedics Phoniatrics Vocology* 33(3), 136-142.
• Deliyski, D.D. (2007). Clinical feasibility of high-speed videoendoscopy. *Voice and Voice Disorders*, 17(1):12-16.
• 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.
• Shaw, H.S. &. Deliyski, D.D. (2008). Mucosal wave: A normophonic study across visualization techniques. *Journal of Voice,* 22(1), 22-33.