Multiparametric Evaluation of Dysphonic Severity

*Estella P.-M. Ma and †Edwin M.-L. Yiu

*Queensland, Australia, and †Hong Kong

Summary: In recent years, the multiparametric approach for evaluating perceptual rating of voice quality has been advocated. This study evaluates the accuracy of predicting perceived overall severity of voice quality with a minimal set of aerodynamic, voice range profile (phonetogram), and acoustic perturbation measures. One hundred and twelve dysphonic persons (93 women and 19 men) with laryngeal pathologies and 41 normal controls (35 women and six men) with normal voices participated in this study. Perceptual severity judgement was carried out by four listeners rating the G (overall grade) parameter of the GRBAS scale.[1] The minimal set of instrumental measures was selected based on the ability of the measure to discriminate between dysphonic and normal voices, and to attain at least a moderate correlation with perceived overall severity. Results indicated that perceived overall severity was best described by maximum phonation time of sustained /a/, peak intraoral pressure of the consonant-vowel /pi/ strings production, voice range profile area, and acoustic jitter. Direct-entry discriminant function analysis revealed that these four voice measures in combination correctly predicted 67.3% of perceived overall severity levels.

Key Words: Voice assessment—Multiple measures—Dysphonic severity—Aerodynamic—Voice range profile—Acoustic perturbation.

INTRODUCTION

Contemporarily, dysphonic severity is evaluated by perceptual judgment and instrumental measurements. Perceptual voice evaluation is regarded by clinicians and researchers as the ‘‘gold standard’’ for documenting voice impairment severity.[2] Because it involves subjective judgment of voice quality and severity, it is susceptible to various sources of inter- and intralistener variability (see the review by Kreiman et al[2]). The literature has shown that perceptual reliability can be affected by the type of rating scale, the vocal quality, and voice samples to be evaluated; the background and experiences of the listeners; and the provision of external voice references as anchors for the listeners. Previous studies have also shown that variability for ratings of individual voices, which is indicated by the width of the 95% confidence interval, is higher for mild-to-moderately rough voices than for voices at the two endpoints (normal and extremely rough) on the rating scale.[2–4]

Accepted for publication April 5, 2005.

Presented at the 33rd Annual Voice Foundation Symposium: Care of the Professional Voice, June 2004, Philadelphia, Pennsylvania.

From the *Division of Speech Pathology, The University of Queensland, Queensland, Australia; and the †Voice Research Laboratory, Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong.

Address correspondence and reprint requests to Estella P.-M. Ma, Division of Speech Pathology, The University of Queensland, St. Lucia, Brisbane, QLD 4072 Australia. E-mail: estella.ma@shrs.uq.edu.au

Journal of Voice, Vol. 20, No. 3, pp. 380–390
0892-1997/$32.00

doi:10.1016/j.jvoice.2005.04.007

Instrumental measurements, on the other hand, frequently involve instrumentation to quantify dysphonic severity. They are regarded as less subjective and hence a more reliable method to document vocal dysfunction. It is therefore not surprising to find the extensive literature identifying which instrumental measure can best predict perceptual severity, with the intention of replacing perceptual evaluation to document voice impairment severity. However, there has been an inconclusive finding of any single instrumental measure can consistently correlate strongly with perceptual judgment. Some researchers considered the multidimensional nature of voice and advocated the use of more than one type of instrumental measure to predict perceptual severity. This multiparametric approach allows simultaneous inclusion of different instrumental voice measures and therefore enhances the power in differentiating perceptual severity levels.[5]

Several authors have investigated the effectiveness of combining different instrumental measures to describe perceptual severity.[5–9] Such effectiveness is commonly evaluated in terms of the association (or concordance) between voice severity levels perceptually judged by listeners and predicted by instrumented measures of the same voice samples. The higher the concordance, the stronger the association between perceptual evaluation and instrumental measurements. Two different statistics have been included in these studies to evaluate concordance between perceptual and instrumental analysis. The most common is agreement; that is, the percentage of voice samples whose severity levels measured by perceptual and instrumental analysis are the same.[5,6,9] Other statistics included correlation coefficient (Pearson's r) between the results of perceptual and instrumental analysis.[8]

Giovanni et al[9] employed two acoustic perturbation (jitter and signal-to-noise ratio) with two aerodynamic (voice onset time and glottal leakage) measures that were collected simultaneously with the EVA® system to predict perceptual severity ratings. Perceptual judgment was performed on a 5-point rating scale from "0" (normal) to "4" (severe). Direct-entry discriminant function analysis revealed that the four instrumental measures in combination achieved 66.1% (158 out of 239) concordance with perceptual severities. However, this concordance was based on voice samples perceptually rated as "0 (normal)," "2 (moderate)," "3 (intermediate)," and "4 (severe)." Voice samples rated as "1 (very light or intermittent voice abnormalities)" were not included in the analysis because these samples did not show significant differences from grade "0" and "2" voice samples. In other words, mildly impaired voice quality was not easily discriminated by the set of acoustic and aerodynamic measures.

Piccirillo et al[7,8] carried out two studies in an attempt to develop a multiparametric voice function index to describe dysphonic severity. They employed the multivariate logistic regression technique and identified a minimal set of four among 14 voice measures that could best distinguish between dysphonic and normal voices. The measures selected were estimated subglottal pressure, phonational frequency range, airflow rate measured at the lips, and maximum phonation time. However, the correlation between the combination of four measures and perceived overall severity was only moderate (Pearson's $r = 0.58$).

Wuyts et al[5] devised the Dysphonic Severity Index with four out of 13 aerodynamic, voice range profile, and acoustic perturbation measures. The four voice measures were statistically selected with the stepwise logistic regression procedure and represented the minimal set of instrumental measures that could best predict perceptual severity. These four measures were jitter percent, maximum phonation time of sustained /a/, the highest frequency value, and the minimum intensity level. Perceptual evaluation was performed on a 4-point scale and was taken from the grade component of the GRBAS scale.[1] However, an integration of these four measures achieved only 49.9% (193 out of 387 subjects) concordance with perceived overall severity.

Yu et al[6] obtained 11 aerodynamic and acoustic perturbation measures with the EVA® system. Perceptual severities were taken from the overall grade of the GRBAS scheme. The authors employed stepwise discriminant function analysis and identified a set of six measures that could most clearly distinguish among perceptual severity