Auditory Perceptual Voice Evaluation

Voice quality

- A wide variety of descriptive characteristics have been utilized over the years to encapsulate how a voice sounds, such as breathy, rough, strained, harsh, metallic, constricted, strident, strained, nasal, denasal, and husky. Phonation breaks such as in laryngeal tremor and spasmodic dysphonia added further challenges to evaluation of voice. The multitude of potential approaches led to a complex landscape of descriptors and measurements, rendering comparison and interpretation difficult. Therefore it was proposed that systematic measurement of disordered voice characteristics would be definable, measurable, independent, and valid [1].

- A complete perceptual evaluation of voice may also involve elements beyond the audio-perceptual dimension of voice quality, such as patterns in how a patient is breathing and visible physical tension / posture, but these components are beyond the scope of the present discussion.

GRBAS

- The GRBAS scale was perhaps the first widely utilized measure, and is still in use today. This includes 5 components: G(rade) -- overall grade of hoarseness, R(oughness), B(reathiness), A(esthenia) -- weakness, and S(train). Each component is rated on an integer four point scale, in which 0 is normal, 1 slight, 2 moderate, and 3 severe. This scale has been well-received because it is brief and user-friendly, making it practical for application in a clinical setting. Further, the presentation of a rating such as $G_2R_2B_1A_1S_2$ allowed for efficient communication of assessments across raters [2].

- Limitations of the GRBAS scale include (1) limited granularity of ratings given the 4 point Likert scale rating options, and (2) challenges of inter- and intra-rater reliability, particularly for abnormal voices; to some degree these challenges can be mitigated by rater training [3].

CAPE-V

- The Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) scale was developed through the American Speech-Language-Hearing Association to be used for standardized perceptual assessment of voice using a psychometrically valid approach. This scale is used with standardized reading content and includes ratings on a 100 mm visual analogue scale, with severity anchors, of overall severity, roughness, breathiness, strain, pitch, loudness, and also includes the options to comment about resonance and other characteristics such as fry, pitch instability, and tremor [4].
- When compared to GRBAS, CAPE-V showed slightly better rater reliability [5].

- Challenges in the use of the CAPE-V can be categorized into signal-based (e.g., how best to rate a voice that is both strained and breathy), rater-based (e.g., unstable internal

standards of voice quality may lead to rating drift, particularly in less experienced listeners), and task-based effects (e.g., raters may have different interpretations of the scale instructions). Those using the CAPE-V should also be aware that two versions exist: one with evenly spaced severity anchors, and another with non-linearly placed severity anchors, and be sure to use the same version if making comparisons based on CAPE-V ratings. Research on optimizing CAPE-V use is ongoing, particularly with respect to reliability and training effects [6].

The future

- Other scales have been proposed in addition to GRBAS and CAPE-V, which were highlighted here due to their broad use. However, because it has proven to be very difficult for listeners to assess single dimensions of voice quality when multiple dimensions are perturbed, newer work has focused on a psychoacoustic model of voice that has identified specific acoustic correlates of how listeners perceive voice quality [7].

- Key acoustic characteristics that most effectively identify differences across individual voices include measures of harmonic source spectral shape, inharmonic source spectral shape, time-varying source characteristics, and vocal tract transfer function. These model parameters are perceptually detectable by listeners. Although further development and validation is ongoing, eventually, subjective perceptual assessment of voice quality using traditional perceptual scales with isolated dimensionality may become unnecessary. Current work is focused on a physiopsychoacoustic model that links voice production, acoustics, and perception [8].

References

1) Perkins WH. Vocal Function: A Behavioral Analysis. In: Travis LE, editor. Handbook of speech pathology and audiology. Prentice-Hall; 1971. p. 481–503.
2) Hirano M. Clinical Examination of Voice. New York, NY: Springer Verlag; 1981, pp. 81-84.
3) Martens JWMAF, Versnel H, Dejonckere PH. The effect of visible speech in the perceptual rating of pathological voices. Arch Otolaryngol Head Neck Surg 2007;133:178–85.
4) Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol. Am J Speech Lang Pathol 2009;18:124–32.
5) Zraick RI, Kempster GB, Connor NP, et al. Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). Am J Speech Lang Pathol 2011;20(1):14–22.
6) Nagle KF. Emerging Scientist: Challenges to CAPE-V as a Standard. Perspect ASHA SIGs 2016;1(3):47–53.
7) Kreiman J, Gerratt BR, Garellek M, Samlan R, Zhang Z. Toward a unified theory of voice production and perception. Loquens 2014;1(1).
8) Kreiman J, Gerratt B. Reconsidering the Nature of Voice. In: Frühholz S, Belin P, editors. The Oxford Handbook of Voice Perception. Oxford University Press; 2019.