Pervasive Diagnosis and Rehabilitation of Voice Disorders: Current Status and Future Directions

Eva van Leer, PhD, MFA, CCC-SLP
Georgia State University
Atlanta, GA 30303
eva.van.leer@gmail.com
evanleer@gsu.edu

Nick Porcaro
Chief Scientist, moForte, Inc
nick@ccrma.stanford.edu

ABSTRACT

Voice disorders affect approximately 7.5 million Americans and more individuals worldwide, have multiple etiologies, and occur throughout the lifespan. Occupational voice users such as teachers and call-center workers are particularly at risk for the development of chronic voice problems. The American Academy of Otolaryngology recommends behavioral voice therapy for a variety of voice disorders. However, approximately 50% of voice patients drop out or experience difficulty practicing and generalizing healthy speaking-voice technique. Recently, the development of mobile tools has been initiated to assist and motivate voice patients between therapy sessions. Such tools provide reminders, voice exercise videos or audio examples, and interactive acoustic analysis software. In addition, several iOS apps have been developed as low-cost alternatives to voice quality analysis suites for objective acoustic measurement of disordered voices. Moreover, singers and other intensive voice users have started the emerging trend of taking advantage of smartphones' camera feature to visualize and video-record their own vocal folds, hereto impossible outside of the medical or research-laboratory environment. This paper introduces the reader to the evaluation and treatment of voice disorders, the current technologies that may enhance or disrupt the practice of voice care, and future directions.

Keywords

Voice Disorders, Voice therapy, treatment adherence, self-efficacy, mobile health

Categories and Subject Descriptors

Human Computer Interaction, Digital Signal Processing

1. INTRODUCTION

Voice disorders have significant and increasing societal impact. Voice disorders are changes in vocal quality, pitch, loudness, effort, and endurance [Ramig & Verdolini 1998]. These affect approximately 7.5 million Americans, (http://www.nidcd.nih.gov/health/statistics/vsl/Pages/stats.aspx; page accessed 03-26-2016), have multiple etiologies, and occur throughout the lifespan, with a prevalence of up to 47% [Roy et al, 2004]. Vocal impairments that are experienced as an inconvenience in an episode of laryngitis become socially and professionally devastating when chronic. Involving the majority of life-participation categories identified by the World Health Organization [International Classification of Functioning, Disability and Health, WHO, 2001], voice disorders negatively affect mental health, increase health care costs and commonly result in short- and long-term disability [Cohen, 2010; Cohen et al 2012]. In today's service-oriented work force, 25-35% of individuals depend primarily on their voice for their occupation [Titze et al, 1997]. Because service-oriented industries are projected to grow, the number of occupational voice users is likely to rise accordingly (http://www.bls.gov/emp/ page accessed 03/26/2016). Therefore, research into improving the efficacy and efficiency of voice treatment is an important physical- and mental-health initiative that holds economic and social consequences. Moreover, maximizing adherence to an existing efficacious treatment is consistent with the goals of the Affordable Care Act [Kocher et al, 2010].

2. BACKGROUND

Voice care is best provided by team of health care professionals that is at a minimum composed of a Laryngologist (i.e. an Ear Nose and Throat physician with further specialization in surgical and medical treatment of voice disorders) and a Speech-Language Pathologist with additional training in the evaluation and treatment of voice disorders [Cohen et al 2012]. Voice evaluation typically includes "stroboscopy:" the visualization of the larynx via a rigid scope attached to a camera, under strobe light. During voice production (i.e. "phonation") the vocal folds complete cycles of oscillation that can range from approximately 80 Hz (a low pitch produced by a male adult) to 1000 Hz (a high note produced by a female adult soprano singer). As such, vocal fold oscillation appears as a blur or hummingbird wing to the human eye. In order to achieve a slow-motion appearance that allows the viewer to examine the structure and function of the vocal folds during oscillation, a strobe light is set to flash at a rate synchronized to the vibratory rate of the vocal folds, but at a slight delay such that consecutive moments in the vibratory cycle are captured. Phonation can then be examined for mechanical parameters such as stiffness, pliability, amplitude of vibration, completeness or pattern of closure when the vocal folds approximate during the closing phase of oscillation, symmetry, and other aspects. At rest, the larynx can be assessed for the presence of any pathology or abnormal changes in color and

PervasiveHealth 2016, May 16-19, Cancun, Mexico

DOI 10.4108/eai.16-5-2016.2264713

structure such as vocal fold nodules, polyps, hemorrhages (i.e. bruising), or swelling.

In addition to visualization of the larynx, voice evaluation may include aerodynamic measures taken during voice production, patient questionnaires that describe the voice problem and its effect on quality of life, and perceptual and objective acoustic measurement of the voice signal. Objective acoustic measures of voice have recently undergone improvements in algorithm or have been combined through regression analysis, such that they better predict perceptual human ratings of voice quality such as overall clarity [Maryn & Weenink 2014]. These measures, such as Cepstral Peak Prominence, have a reasonably good (i.e. above .8) correlation with human perception of voice quality [Awan, Roy, & Dromey 2009] and may therefore hold promise to be utilized to assist patients in tracking their own voice quality objectively.

For a variety of organic, neurogenic and behavioral voice disorders, behavioral voice therapy is recommended in the Clinical Practice Guideline of the American Academy of Otolaryngology- Head & Neck Surgery [Schwartz et al 2009]. In voice therapy, patients learn to alter their voice production mechanics or technique from their habitual damaging or fatiguing approach to one that optimizes voice quality and resilience while reducing or resolving pathology [Thomas and Stemple, 2007]. A target technique (e.g. resonant voice, confidential voice, breathy voice) is acquired through direct instruction during voice therapy sessions, and further depends on independent patient practice and use in connected speech in all communication contexts outside of weekly therapy sessions [Ramig and Verdolini 1998]. Achieving consistent healthy voice production, therefore, involves motor learning and extensive self-regulation [Wulf et al 2010]: the volitional goal-directed planning, self-evaluation, and self-correction of a new behavior.

Although voice therapy is effective, adherence is a challenge for patients, as only 35% of patients complete therapy [Hapner et al, 2009; Portone et al, 2008].The 65% drop-out rate is comparable to that of health behavior programs such as physical exercise, diabetes self-management, and smoking cessation [Glanz et al 1990; DiMatteo et al 2002]. Moreover, those patients who attain voice therapy goals often deteriorate after treatment completion [van Lierde et al 2007], indicating poor long-term maintenance of healthy voice use. Resolving this adherence problem is imperative to maximizing treatment efficacy for all individuals with voice disorders- not only those who are innately skillful in technique change.

3. PERVASIVE VOICE EVALUATION

Both the evaluation and treatment of individuals with voice disorders can be improved through pervasive technologies. Voice evaluation can be made more accessible via low-cost and off-the-shelf technologies that allow individuals with voice disorders to self-assess their vocal health or, at a minimum, provide relevant information to their health care provider remotely. Just as one can measures one's own glucose levels anytime, anywhere, for self-care and communication with the health care provider, one should be able to track aspects of one's vocal fold health accordingly. Greater patient involvement in voice evaluation would represent a dramatic disruptive shift in the current model of health care, wherein voice evaluation requires a clinic visit due to patients' inability to visualize their own vocal folds. Subjective self-report (e.g. "I'm hoarse today: I wonder if my vocal folds are swollen") is limited in informational value compared to objective data in the form of vocal fold images. Thus, development of pervasive self-evaluation tools could reduce the gap between patient and provider.

Along these lines, a disruptive popular trend has been noted in which individuals are using their smart phone to visualize their vocal folds themselves outside of the medical setting, as noted here: https://www.youtube.com/watch?v=5_RbH3uK640. Individuals' success in doing so is in part dependent on the width of the phone. A narrower phone is more likely to allow for vocal fold visualization, as the size of the oral cavity is limited, and the vocal folds (in the neck) are examined from the posterior oral cavity. The further development of existing commercial mobile phone telescoping lenses to better suit the vocal tract may aid in successful laryngeal imaging outside of the clinic. This will allow individuals to examine their vocal folds at rest, and assess ability of the vocal folds to move medially to the center as well. Strobe apps that can flash at several rates, in synchrony with an audio signal or "flash to the beat of the music" (e.g. Strobily) as well as those that flash during video recording hold the potential to also examine vocal fold oscillation. Because voice care is a specialization within Otolaryngology and Speech-Language pathology, specialty voice care is limited primarily to centers within larger cities. Performers who rely on their voice are frequently travelling and develop illness or vocal trauma on the road. Patient self-imaging of the vocal folds, and stroboscopic imaging can provide a potential communication route to their established voice provider. This would be particularly beneficial in the case of vocal fold hemorrhaging, where complete voice rest (i.e. silence) is the recommended until the hemorrhage is resolved. Without access to laryngeal imaging, performers typically worsen vocal injury by continuing to perform or talk in the presence of hemorrhage or vocal fold inflammation. Thus, development of hardware adds (e.g. telescoping lenses) and apps can alter and improve the way we care for travelling voices.

Objective acoustic analysis of voice is a second diagnostic area in which pervasive technology can play a substantial role. Although acoustic measures cannot provide a diagnosis- laryngeal visualization is necessary for diagnosis- objective measurement of voice can help track vocal improvement or deterioration. Traditional systems for acoustic analysis have previously typically required costly hardware and software, although freeware (e.g. PRAAT) is also available. For voice analysis, iOS platforms can accommodate a high sampling rate (i.e., 32–44 KHz compared to the telephone's 8 KHz), 24 bit analog-to-digital conversion, a flat frequency response between 20–20,000 KHz (if one disables the filter attenuating signals below 200 Hz), and uncompressed audio file format (.wav). By recording via a microphone with a documented flat frequency response directly into an iOS device in an uncompressed file format (i.e. .wav), perturbation, spectral, and cepstral measures of voice may be calculated locally, as iOS devices hold adequate computing power and vary <5 dB between devices. This option can reduce cost and improve mobility for travelling voice clinicians who, with specialty expertise, provide voice care at multiple locations. One example of an iOS based acoustics lab that is currently available is OperaVox (OperaVox LLC), which calculates several voice parameters including fundamental frequency and perturbation measures. Furthermore, reliable commercially available intensity applications include Soundmeter (Faber Acoustics) and AudioTools (StudioSix). Future tools might incorporate additional spectral and cepstral measures for comprehensive analysis.

4. PERVASIVE VOICE THERAPY TOOLS

Patients report difficulty in accurately replicating a target voice production technique independently without clinician feedback [van Leer and Connor, 2010]. Among other barriers to adherence, they have difficulty remembering to practice and use the technique throughout the day in conversation, in part because this requires self-regulatory effort, and in part because the problematic baseline technique is well established. Therefore, to improve short- and long-term efficacy of voice therapy, adherence tools must be developed that help patients accurately reproduce the target voice technique independently as well as help patients remember and implement practice and generalize.

Telemedicine technologies can allow for communication between clinician and patient, but require clinician time. Clinicians cannot be available at all times that voice production feedback is needed. Moreover, many aspects of voice therapy adherence either do not require clinician expertise (e.g. scheduling daily practice and remembering to practice) or require consistent support (e.g. self-monitoring voice quality during a conversation). Unlike clinicians, mobile solutions are available anytime, anywhere, and can support self-directed patient learning in vocal rehabilitation.

The development of pervasive technologies to assist voice patients in extra-clinical practice and generalization is significant because these tools have the potential to 1) directly improve adherence: the primary limiting factor in voice rehabilitation, 2) inform our theoretical and empirical understanding of motor learning, goal-directed behavior change, and communication 3) have utility for use in both the general population and for unique populations such as choral singers and pastors, as well as the deaf and hard-of-hearing.

Existing functionalities of mobile platforms can address the barrier of remembering to practice. An abundance of calendar functions and to-do-list applications exist to assist the patient. What is currently a more challenging problem to solve is assistance with the self-directed learning in which patients engage between formal therapy sessions.

A theory-driven approach to improving frequency and accuracy of home practice is to provide both external support to directly influence practice behavior, and to influence practice indirectly by improving motivation for practice. Specifically, Social Cognitive Theory's model of Triadic Asymmetrical Reciprocal Causation speaks to the importance of self-efficacy for practice [Bandura 1986]. Although perceived vocal handicap may motivate a person to seek medical care for the voice, patient-perceived severity of the disorder has not shown to predict adherence to behavioral voice treatment [Smith et al 2010, van Leer & Connor 2015]. Thus, an individual may be highly distraught regarding the gradual loss of speaking-voice or inability to sing, but this does not have bearing on vocal behavior change or practice of voice exercises. Rather, in our work, self-efficacy (i.e. task-specific confidence) for practice of voice technique and generalization, and the quality of the therapeutic relationship between patient and clinician, were found to be predictors of adherence.

One approach to improving practice accuracy is the provision of video models. Existing video cameras on patients' smartphones can record important segments of voice technique instruction during the in-person session for future review by the patient. Video models may be more useful than audio recordings only, as video provides relevant information such as posture, relaxation of the jaw and lips, and if in frame, movement of the ribcage and abdominal wall. There is extensive evidence to support the use of video examples, in particular video-self-models, for a variety of behaviors as well as for adherence to fluent speech in the treatment of stuttering [Cream et al 2009]. In our previous experimental studies, clinician and patient video examples on MP4 players were found to significantly increase self-reported patient adherence and self-efficacy for adherence [van Leer, 2010; van Leer & Connor, 2012). In a randomized cross-over trial, patients practiced twice when provided with MP4 players containing therapy video examples, than when practicing with written homework instructions. Self-efficacy for practice was also significantly increased in the MP4 condition. Thus, both patients' actual practice frequency and their confidence or motivation for practice were improved with mobile video examples. In a subsequent randomized trial, practice frequency in the MP4 group was not significantly greater than for individuals in the control condition, but participants in the MP4 group reached a significantly higher level of generalization by the end of the study, and higher voice outcome values. In this study then, it appears that when provided with MP4 video examples of therapy, patients move through the motor skill acquisition phase and onto the generalization phase more quickly. The initial period of confusion between sessions may have been reduced by concrete video examples in this stage of learning. As predicted, video self-models (in which the patient was recorded using the target voice technique) were a source of self-efficacy: "I saw that I can do this; and it wasn't weird: I didn't look weird." Mobile video support was particularly beneficial to those with initially low self-efficacy for voice therapy.

An unexpected finding was that patients in the MP4 group rated the quality of the therapeutic relationship higher than patients in the control group. Patients stated that clinician video models not only helped them recall how to produce the target voice but also that "having the clinician with me all the time" in video format felt supportive. Thus, the therapeutic relationship was improved without training clinicians in relationship-building skills.

Although video models exemplify the target technique and help patients recall the technique, they are not interactive and therefore cannot correct poor practice. Extrinsic feedback is fundamental to motor learning; in fact, no learning can occur without feedback [Wulf & Shea 2010]. Extrinsic feedback provided at the learner's request (i.e. self-controlled feedback) after completion of a motor task and regarding the effect or end result of a movement facilitates motor learning and increases motivation, in turn facilitating further learning. Extrinsic feedback is essential to self-regulation and motor skill acquisition by allowing individuals to note the discrepancy between their current performance and the performance goal. Feedback also improves self-efficacy and emotional affect, and thus motivates continued effort toward goal attainment. Therefore, it is reasonable to assume that provision of voice quality feedback between sessions would increase skill acquisition and patient motivation to attain voice therapy goals.

Objective acoustic analysis methods can potentially do so in the clinician's absence by capturing and displaying information about aspects of patient voice quality. The challenge of developing feedback tools lies not in limitations of mobile technology, but in the state-of-voice-science itself. Most easily computed are the objective measures of frequency and intensity measures. These correspond to the perceptual aspects of vocal pitch and loudness. Although pitch and loudness are important aspects of voice, they are typically only therapy targets for a limited population. For example, individuals with dysphonia related to Parkinson's disease undergo active training to increase vocal loudness, which also has positive effects on articulation in this particular population. Likewise, individuals with vocal fold lesions due to excessive loud strain may benefit from continuous feedback regarding loudness level in order to reduce loudness. Those with the goal to reduce the use of "vocal fry," a rough-sounding low-pitched voice production, can benefit from feedback regarding frequency. Commercially available intensity applications include Soundmeter (Faber Acoustics) and AudioTools (Studio Six); real time pitch feedback is provided by the Passaggio 1.3 app (WIV LLC) whereas EVA Pitch (Exceptional Voice) gives interactive pitch feedback (bug fixes are in progress). Wearable devices can provide the user with pitch and loudness information for improved self-monitoring [van Stan et al 2015].

Unlike pitch and loudness, common voice therapy targets represent *timbre* or *voice quality* parameters, such as "resonant" or "breathy" voice. Objective measurement of voice quality remains a challenge in voice science, in particular in the context of connected speech and noisy naturalistic environments. In particular, measurement of voice quality in connected speech (i.e. at the word or sentence level) is more complex than in sustained phonation (i.e. holding a note). In the author's experience patients' observation of a real-time power spectrum can assist their reaching a voice quality target, as changes in voice quality have spectral consequences. Voice Analyzer [Dexus] is one example of a commercially available spectral application that may be usable by patients without excessive training.

Cepstral measures, in particular cepstral peak prominence (CPP), are associated perceptual ratings of voice quality [Awan, Roy, & Dromey 2009]. CPP highly correlates with human perception of voice quality and is sensitive to treatment-related improvement for both sustained phonation and connected speech. In our laboratory, we have developed and are testing an app that, among other functionalities, calculates CPP of a sound sample. The functionality can be used both for sustained phonation ("aaaah") and connected speech (e.g. "my oh my oh my oh my!"). In a voice therapy session, the patient can measure CPP when producing voice in their habitual fashion as a baseline measure. Once a target voice technique has been established with the clinician, CPP can be measured again. This value can be used as concrete goal for replication at home. Since patients have difficulty judging whether they are "in the good voice or not," this objective measure can assist.

In a usability study of mobile CPP provision [van Leer et al 2016], patients found numeric CPP values useful in making the attainment of the target voice more "concrete." Individuals with low self-efficacy for unassisted practice had significantly greater self-efficacy for practice during practice with CPP feedback. However, unexpectedly, even those patients with high self-efficacy enjoyed CPP feedback because it made practice "more fun." Several patients reported the desire to compete with themselves to obtain a higher CPP value. Thus, there is a potential for gamification in voice practice. The desire to compete with oneself is exploited in the gamification of learning [Deterding et al., 2013], holding great promise for the development of voice therapy games.

5. NEXT STEPS

The introduction of Apple's Research Kit and Care Kit [http://www.apple.com/researchkit/] can increase data collection regarding vocal rehabilitation and habilitation beyond the current number of study participants personally recruited in an investigator's laboratory and affiliated clinics. Voice diagnosis and treatment studies usually employ a small N (e.g. 20-60 individuals) because dedicated voice centers are few and the population of individuals with voice disorders is not as large as common chronic illnesses such as COPD, diabetes, or obesity. Increasing the number of study participants via Research Kit can allow data collection on previously unreachable large samples of patients, thus increasing effect size and reducing spurious findings. Voice recordings, statistics and survey data can be communicated to the laboratory securely by importing data directly to the cloud, without requiring study participants to visit the lab or clinic.

Mobile voice feedback tools can incorporate dimensions that describe the user's voice comprehensively and provide individualized feedback. Not only can current therapy be enhanced, but maintenance of treatment gains can be supported. Furthermore, age-specific games, feedback and tracking mechanisms can be developed for individualization across the lifespan. Pediatric voice therapy adherence is under study by the author and collaborator Maia Braden at the University of Wisconsin, but otherwise receives little empirical investigation. Furthermore, presbyphonia- dysphonia associated with advanced aging- is treatable with voice therapy, and can potentially be prevented via voice exercises largely guided by telepractice and mobile applications. In all, we hope to move from an episodic, location-specific model of voice care to a continuous, integrated model. Given the behavioral self-care and self-directed learning required of voice patients, it is entirely appropriate to empower them with tools for success.

REFERENCES

[1] Awan SN, Roy N, Dromey C. Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model.Clin Linguist Phon 23:825-841.

[2] Bandura, A. Social foundations of thought and action: A social cognitive theory. Prentice-Hall series in social learning theory. Prentice Hall, Englewood Cliffs New Jersey, 1986.

[3] Cohen SM. Self-reported impact of dysphonia in a primary care population: An epidemiological study. Laryngoscope 120:2022-2032, 2010.

[4] Cohen SM, Kim J, Roy N, Asche C, Courey M. The impact of laryngeal disorders on work-related dysfunction. Laryngoscope 122:1589-1594, 2012.

[5] Cream, A., O'Brian, S., Onslow, M., Packman, A., & Menzies, R. (2009). Self-modelling as a relapse intervention following speech-restructuring treatment for stuttering. International Journal of Language & Communication, 44(5), 587-599.

[6] Deterding s, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification." MindTrek conference proceedings, 11 9-15, 2011.

[7] DiMatteo M, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. Med Care 40:794-811, 2002.

[8] Glanz K, Rimer BK, Marcus Lewis F. Theory, research and practice in health behavior and health education. In: Glanz K, Rimer BK, Marcus Lewis F, eds. Health Behavior and Health Education: Theory, Research and Practice. 3rd ed. San-Francisco: Jossey-Bass; 2002:34.

[9] Hapner E, Portone-Maira C, Johns III MM. A study of voice therapy dropout. J Voice 23:337-340, 2009.

[10] Kocher R, Emanuel EJ, DeParle NM. The Affordable Care Act and the future of clinical medicine: the opportunities and challenges. Ann Intern Med 153:536-539, 2010.

[11] Portone C, Johns MM,3rd, Hapner ER. A review of patient adherence to the recommendation for voice therapy. J Voice 22:192-196, 2008.

[12] Ramig LO, Verdolini K. Treatment efficacy: voice disorders. J Speech Lang Hear Res 41:S101-116, 1998.

[13] Roy N, Merrill RM, Thibeault S, Parsa RA, Gray SD, Smith EM. Prevalence of voice disorders in teachers and the general population. J Speech Lang Hear Res 47:281-293, 2004. Rechenberg L, Goulart BN,

[14] Schwartz SR, Cohen SM, Dailey SH, Rosenfeld RM, Deutsch ES, Gillespie MB, Granieri E, Hapner ER, Kimball CE, Krouse HJ, McMurray JS, Medina S, O'Brien K, Ouellette DR, Messinger-Rapport BJ, Stachler RJ, Strode S, Thompson DM, Stemple JC, Willging JP, Cowley T, McCoy S, Bernad PG, Patel MM. Clinical practice guideline: Hoarseness (Dysphonia). Otolaryngol Head Neck Surg 141:S1-S31, 2009.

[15] Smith BE, Kempster GB, Sims HS. Patient factors related to voice therapy attendance and outcomes. J Voice, 24: 697-701, 2010

[16] Thomas LT, Stemple JC. Voice therapy: Does science support the art? Comm Dis Rev 1:49-77, 2007.

[17] Titze IR, Lemke J, Montequin D. Populations in the U.S. workforce who rely on voice as a primary tool of trade: a preliminary report. J Voice 11:254-259, 1997.

[18] van Leer E, Connor NP. Patient perceptions of voice therapy adherence. J Voice 24:458-469, 2010.

[19] van Leer E, Connor NP. Use of portable digital media players increases patient motivation and practice in voice therapy. J Voice 26:447-453, 2012.

[20] Van Leer, E, Connor N.P. Predicting and Influencing Patient Adherence using Social-Cognitive Factors and Mobile Video. American Journal of Speech-Language Pathology, May 2015, Vol. 24, 164-176.

[21] van Leer E, Hapner ER, Connor NP. Transtheoretical Model of Health Behavior Change Applied to Voice Therapy. J Voice 22:688-698, 2008.

[22] Van Leer, Eva, Robert C. Pfister, and Xuefu Zhou. "An iOS-based Cepstral Peak Prominence Application: Feasibility for Patient Practice of Resonant Voice." Journal of Voice (2016).

[23] Van Lierde KM, Claeys S, De Bodt M, van Cauwenberge P. Long-term outcome of hyperfunctional voice disorders based on a multiparameter approach. J Voice 21:179-188, 2007.

[24] Van Stan, Jarrad H., Daryush D. Mehta, and Robert E. Hillman. "The effect of voice ambulatory biofeedback on the daily performance and retention of a modified vocal motor behavior in participants with normal voices." Journal of Speech, Language, and Hearing Research 58.3 (2015): 713-721

[25] Wulf G, Shea C, Lewthwaite R. Motor skill learning and performance: a review of influential factors. Med Educ 44:75-84, 2010.