INTERSPEECH 2017 TUTORIAL PROPOSAL

Title

Real-world ambulatory monitoring of vocal behavior

Presenter

Daryush D. Mehta, PhD

Contact Information

Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital
One Bowdoin Square, 11th Floor
Boston, Massachusetts 02114 (USA)
617-599-0328 (cell), 617-643-2466 (work)
mehta.daryush@mgh.harvard.edu

Abstract

Many of us often take verbal communication for granted. Individuals suffering from voice disorders experience significant communication disabilities with far-reaching social, professional, and personal consequences. This tutorial provides an overview of long-term, ambulatory monitoring of daily voice use and in-depth discussions of interdisciplinary research spanning biomedical technology, signal processing, machine learning, and clinical voice assessment. Innovations in mobile and wearable sensor technologies continue to aid in the quantification of vocal behavior that can be used to provide real-time monitoring and biofeedback to facilitate the prevention, diagnosis, and treatment of behaviorally based voice disorders.

Description

Real-world monitoring of body system functions continues to receive growing interest as *mobile and wearable devices* become more ubiquitous while individuals engage in their normal daily activities. One area of active research has been the development of biomedical devices for the ambulatory monitoring of daily voice use. In particular, these types of devices can play an important role for medical applications to help facilitate the diagnosis and management of vocal abnormalities that may not be readily apparent during routine clinical examinations. Advancements in mobile and wearable technologies continue to enhance our ability to monitor voice and speech characteristics in real-world environments for the improved assessment and treatment of behaviorally based disorders.

Because vocal behavior and its characteristics during real-world situated interaction are considered to play major roles in the etiology of many common voice disorders, clinicians focus a great deal of attention on attempting to *evaluate and modify how patients actually use their voices*. Such efforts are often limited by a reliance on patient self-reporting and self-monitoring. Given the variance inherent in perceptual judgments, these approaches are very subjective and likely to be unreliable. Furthermore, patterns of voice use and misuse typically become highly habituated and may be carried out below an individual's level of consciousness. Thus, objective measures related to the actual role of daily voice use in the etiology of voice disorders have been developed using wearable sensors aid in quantifying what constitutes normal levels of daily voice use.

This tutorial will provide attendees with an in-depth discussion of 1) the evolution of ambulatory voice monitoring technologies, 2) algorithms for processing long-term recordings of voice characteristics, 3) physics-based modeling of ambulatory signals, and 4) clinical and medical applications.

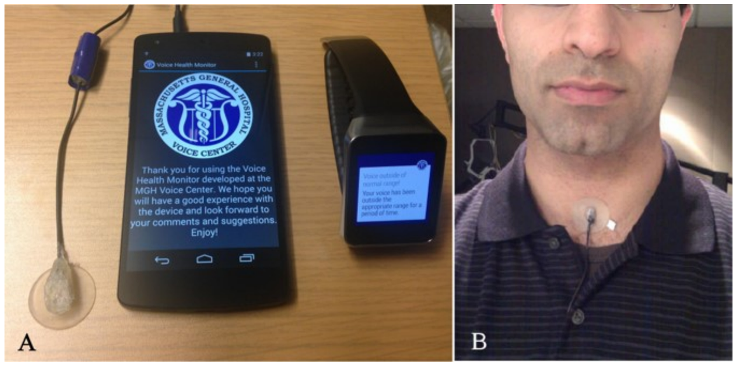

Figure 1. Smartphone-based ambulatory voice monitor: (A) accelerometer, interface cable, smartphone, smartwatch; (B) accelerometer mounted on a silicone pad affixed to the neck above the collarbone.

The general concept of mobile monitoring of daily voice use motivated several *early attempts to develop speech and voice accumulators* that simply used head-mounted acoustic microphones or neck-mounted contact microphones to keep track of the total accumulated duration of voicing. As technologies have advanced, novel wired and wireless ambulatory sensors have been designed, including high-bandwidth accelerometers that, when positioned appropriately, record voice source information that is robust to acoustic environmental noise and alleviates confidentiality concerns associated with recording the acoustic speech signal. These wearable sensors have been paired with mobile digital devices, smartphones, and smartwatches to maximize user compliance and enhance the quality and type of data collected (see Figure 1). Results have verified the modulation of voice use as a function of occupation, which is consistent with observations that individuals in high-voice-use jobs (e.g., singers and teachers) have more frequent voice complaints than those in low-voice-use jobs (e.g., bookkeepers).

Non-acoustic recordings of neck-surface vibration have provided the ability to quantify *ambulatory voice use characteristics in multiple dimensions*. Voice-related parameters for voice disorder classification fall into the following two categories: (1) time-varying trajectories of features that are computed on a frame-by-frame basis and (2) measures of voice use that accumulate frame-based metrics over a given duration (i.e., vocal dose measures). These measures may be computed offline in a post hoc analysis of data or online on mobile devices for real-time display or biofeedback applications. Primary frame-based measures include overall sound pressure level, fundamental frequency, and spectral and cepstral properties. Using these types of frame-based features, vocal dose measures were developed in an effort to quantify the accumulated load on the vocal system. This voice dosimetry framework was motivated by the desire to establish safety thresholds regarding exposure of vocal fold tissue to vibration during voicing, analogous to Occupational Safety and Health Administration guidelines for auditory noise and mechanical vibration exposure.

As with voice dosimetry, recent studies have continued to apply computational modeling that is designed to more accurately reflect pathophysiological processes of voice production. For example, aerodynamic voice measures (related to airflow and pressure) have been derived from neck-surface acceleration using a *biologically inspired acoustic transmission line model* that allows for the estimation of glottal airflow. The model must be individually tuned to anatomical characteristics of individual speakers to optimize the estimation of these measures. Ultimately, the goal of extracting such physiologically salient measures is to provide more relevant inputs to big data processing techniques such as supervised and unsupervised machine learning that can analyze patterns of vocal behavior from large numbers of speakers. Forays into this area will be discussed.

The tutorial wraps up with specific applications of ambulatory voice monitoring and biofeedback technologies for the assessment of speakers in high-voice-use occupations and treatment of individuals with common voice disorders. Disturbances in voice production can be caused by a variety of conditions that affect laryngeal function, including neurological disorders (e.g., Parkinson's disease), congenital or acquired organic disorders (e.g., cancer), and behavioral disorders involving vocal abuse and fatigue. Teachers have been the subject of many research efforts because they work in a high-voice-use occupation and are at a higher-than-normal risk of developing a voice-use-related disorder. Individuals with Parkinson's disease have used voice monitors to help themselves speak louder and communicate more clearly in a crowd. Finally, key application areas in the clinical assessment and treatment of voice disorders will be presented since many common voice disorders are associated with faulty patterns of chronic vocal behavior. Systems integrating smartphones and smartwatches can provide critical biofeedback capabilities through audio, visual, and vibrotactile modalities.

In conclusion, monitoring the vocal behavior of individuals as they go about their daily lives continues to receive attention to quantify healthy speakers and patients with a variety of medical conditions. This tutorial will be designed to introduce the audience to the field, with detailed discussions of hardware technologies, algorithms, and signal processing applications. The proposed tutorial on real-world data collection fits well under the theme of INTERSPEECH 2017 theme of "Situated Interaction" as wearable and mobile device progress for the end goal of interacting with robots and machines.

References

1. J. Gustafsson, S. Ternström, M. Södersten, and E. Schalling, "Motor-learning-based adjustment of ambulatory feedback on vocal loudness for patients with Parkinson's disease," Journal of Voice, vol. 30, no. 4, pp. 407–415, 2016.
2. I. R. Titze and E. J. Hunter, "Comparison of vocal vibration-dose measures for potential-damage risk criteria," Journal of Speech, Language, and Hearing Research, vol. 58, no. 5, pp. 1425–1439, 2015.
3. D. D. Mehta, M. Zañartu, S. W. Feng, H. A. Cheyne II, and R. E. Hillman, "Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform," IEEE Transactions on Biomedical Engineering, vol. 59, no. 11, pp. 3090–3096, 2012.

Relevance

The topic of the proposed tutorial is relevant, timely, and complements the INTERSPEECH 2017 theme of "Situation Interaction." Real-world monitoring of voice and speech characteristics is an active area of research and has the potential to play important roles in the technological advancement of social robots and human-machine interaction. Furthermore, wearable sensors and mobile technologies continue to mature and allow us to seamlessly integrate personal devices in our daily lives. The proposed tutorial also resonates with the typical INTERSPEECH participant by bringing together speech science, signal processing, biomedical sensor technology, and medical applications.

In terms of timeliness, there is growing interest in developing medical devices for monitoring the function of body systems while individuals engage in their normal daily activities (at work, at play, etc.) to help facilitate the detection and diagnosis of abnormalities that may not be readily apparent during routine clinical examinations. Versions of systems exist for ambulatory monitoring of heart, respiratory, gastrointestinal, and brain function. One major impetus for expanding ambulatory monitoring capabilities is the desire to provide remote surveillance of health status and facilitate prevention and timely intervention, particularly as an aging population strives to live independently. Telehealth and telemedicine are hot topics due to pushes for lower medical costs, more personalized clinical treatment, and tracking one's own health.

Clinical approaches to the diagnosis and treatment of voice disorders could be greatly improved by the availability of instrumentation that provides continuous unobtrusive monitoring of important vocal parameters for an extended period of time. Current versions of ambulatory voice monitoring devices have begun to produce valuable new insights into daily patterns of voice use, and enthusiasm remains high for the potential of these devices to improve the diagnosis and treatment of voice disorders. In addition, the application of big data techniques and machine learning approaches has just begun to be applied to this field. Thus, the tutorial may also raise awareness among computer science and machine learning experts for these clinical application areas.

Logistics

Duration: 3 hours (1 session)
Format: One presenter, projector with speakers, hands-on technology demos
Special equipment: None
Accompanying material: Handouts emailed to participants in advance

Presenter Information

Biography
Daryush Mehta is Research Staff (Assistant Investigator) at the Center for Laryngeal Surgery and Voice Rehabilitation at the Massachusetts General Hospital (MGH), Assistant Professor of Surgery at Harvard Medical School, and Adjunct Assistant Professor at the MGH Institute of Health Professions. Daryush received his PhD in Speech and Hearing Bioscience and Technology from the Harvard–MIT Division of Health Sciences and Technology (2010), Master's degree in Electrical Engineering and Computer Science from MIT (2006), and Bachelor's degree in Electrical Engineering from University of Florida (2003). Daryush's research spans voice and speech signal processing, advanced laryngeal imaging, and clinical voice disorder assessment. Read more about his work at scholar.harvard.edu/dmehta.

Key Publications
1. **D. Mehta**, J. Van Stan, and R. Hillman, "Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 4, pp. 659–668, 2016.
2. **D. D. Mehta**, J. H. Van Stan, M. Zañartu, M. Ghassemi, J. V. Guttag, V. M. Espinoza, J. P. Cortés, H. A. Cheyne II, and R. E. Hillman, "Using ambulatory voice monitoring to investigate common voice disorders: Research update," Frontiers in Bioengineering and Biotechnology, vol. 3, no. 155, pp. 1–14, 2015.
3. M. Ghassemi, J. H. Van Stan, **D. D. Mehta**, M. Zañartu, H. A. Cheyne II, R. E. Hillman, and J. V. Guttag, "Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules," IEEE Transactions on Biomedical Engineering, vol. 61, no. 6, pp. 1668–1675, 2014.

Audience Information

Target audience: New researchers to the field, students, specialists in the field
Estimated number of participants: 50
Interested researchers: Sten Ternström, Maria Södersten, Johan Sundberg, Eric Hunter, Thomas Quatieri, Jón Guðnason
Comments: This tutorial will be accessible to researchers from a wide variety of backgrounds, including biomedical devices, speech science and technology, speech-language pathology, linguistics, machine learning, and signal processing.
Bibliography: See Description above.