ICVPB 2018

**11th International Conference on
Voice Physiology and Biomechanics**

August 1-3, 2018, East Lansing, Michigan, USA

2018

Committees

Thank you for all who have spent time and effort preparing for ICVPB 2018!

ICVPB 2018 Local Committee

• Eric Hunter, PhD (Chair)
• Dimitar Deliyski, PhD (Co-Chair)
• Jeff Searl, PhD
• Peter Lapine, PhD
• Maryam Naghibolhosseini, PhD
• Lady Catherine Cantor Cutiva, PhD
• Russ Banks, CCC SLP
• Mark Berardi
• Hamzeh Ghasemzadeh
• David Ford, CCC SLP
• Kim Winkel
• Elizabeth Deliyski
• Andrea Schmitt
• Julia Ann Andary
• Angela Zelnicka
• Melissa Kleinfeld
• Lizzy Gifford
• Samantha Gehm

ICVPB 2018 Scientific Committee

• Eric Hunter, PhD (Chair)
• Dimitar Delliyski, PhD (Co-Chair)
• Jeff Searl, PhD
• Peter Lapine, PhD
• Maryam Naghibolhosseini, PhD
• Lady Catherine Cantor Cutiva, PhD
• Pasquale Bottalico, PhD

ICVPB Advisory Committee

Chair:
• Brad Story, PhD

Members:
• Ingo Titze, PhD
• Johan Sundberg, PhD
• Nathalie Henrich, PhD
• Christine Shadle, PhD
• Paavo Alku, PhD
• Chuck Larson, PhD
• Jan Svec, PhD

Sponsors

Thanks to the following sponsors of ICVPB 2018

comunikrte

College of Communication Arts and Sciences, Michigan State University

Dept of Communicative Sciences and Disorders, Michigan State University

Table of Contents - Proceedings

AUDITORY DISCRIMINATION OF DIFFERENT TYPES OF SONIFIED SYNTHESIZED GLOTTAL AREA WAVEFORMS

Philipp Aichinger[1], Imme Roesner[1], Franz Pernkopf[2], Jean Schoentgen[3]

[1]Division of Phoniatrics-Logopedics, Department of Otorhinolaryngology, Medical University of Vienna, Austria
[2]Signal Processing and Speech Communication Lab, Graz University of Technology, Austria
[3]Department of Bio-, Electro- And Mechanical Systems, Faculty of Applied Sciences, Université Libre de Bruxelles, Belgium

Keywords: Listening Experiment; Glottal Area Waveform; Voice Source Modelling; Voice Quality Typing

INTRODUCTION

Voice quality typing is relevant to the clinical care of voice disorders. It is an important factor in the indication, selection, evaluation, and optimization of clinical treatment techniques. We investigate the human ability to auditorily discriminate different types of sonified synthesized glottal area waveforms (GAW) with a view to closing the knowledge gap that exists between production kinematics and auditory percepts.

METHODS

Test signals

Based on observations made on laryngeal high-speed videos of 120 subjects [1], eight distinct types of GAWs are defined and implemented in a GAW synthesizer. These GAW types are (1) random extra pulses during the closed phases [2], (2) diplophonia [3], (3) fixed phase differences between two cyclic pulse trains, (4) periodic sawtooth-shape modulated phase differences, (5) periodic sinus-shape modulated phase differences, (6) randomly reset sawtooth-shape modulated phase differences, (7) white Gaussian noise modulated phase differences, and (8) bicyclicity. The signals are 2 seconds of duration and the sampling rate is 48 kHz.

The different GAW types are synthesized as follows. Pulse trains are generated with fundamental frequencies $f_o \sim \mathcal{N}(175, 50^2, 50, 600)$ and open quotients $OQ \sim \mathcal{N}(.6, .15^2, .1, 1)$, sampled from truncated Gaussian distributions $\mathcal{N}(\mu, \sigma^2, l, u)$, where μ and σ^2 are the mean and the variance, and l and u are the lower and upper bounds. Hann windows p_h are used for creating pulse shapes $p = p_h^s$ with exponent $s \sim \mathcal{N}(1, .2^2, .5, 1.5)$ serving as a pulse shaping parameter. The amplitude of the pulse trains is 1, unless declared otherwise. The parameters are selected empirically so as to enable synthesis of naturally looking GAWs.

In type (1), a cyclic pulse train is mixed with a random-amplitude pulse train. Random pulses are triggered by means of a Bernoulli distribution, with the extra pulse rate parameter $\rho \sim \mathcal{N}(.3, .1^2, .02, .9)$. Random pulses occur at fixed times between the adjacent cyclic pulses. The amplitude of the cyclic pulses is 1 and the amplitude of the random pulses is $h \sim \mathcal{N}(.3, .05^2, .1, .5)$.

In type (2), two cyclic pulse trains with different fundamental frequencies are mixed. The ratio of the two frequencies $r = a/b + c \cdot \pi/10$, where a, b, and c are random integers drawn from uniform discrete distributions [1,…,9], [2,…,17], and [0,1], respectively. Thus, c switches r from rational to irrational. The fundamental frequencies are in the range [50,600], and $r \neq 1$. The amplitudes of the two pulse trains are $h \sim \mathcal{N}(.3, .05^2, .1, .5)$ and $1 - h$.

In types (3) to (7), two cyclic pulse trains with amplitudes of 0.5 are added together. For type (3), the phase shift $\phi \sim 2 \cdot \pi \cdot \mathcal{N}(.3, .1^2, .1, .49)$ between the two pulse trains is fixed. For types (4) and (5), the phase shift is time-variant. A maximal phase shift $\phi_{max} \sim 2 \cdot \pi \cdot \mathcal{N}(.3, .1^2, .1, .49)$ is obtained, and a meta cycle length of m cycles is drawn from a uniform discrete distribution [2,…,10]. For type (4), the phase shift increases linearly from 0 to ϕ_{max} until it is reset to 0 in the mth cycle. For type (5), the phase shift is sinusoidally modulated over the length of m cycles and with an amplitude of ϕ_{max}. For type (6), the phase shift increases linearly from one cycle to the next in steps of $\phi_\Delta \sim 2 \cdot \pi \cdot \mathcal{N}(.3, .1^2, .1, .49)$. The phase shift is reset to 0 in cycles randomly selected by means of a Bernoulli distribution, with phase shift reset rate $\rho_r \sim \mathcal{N}(.3, .1^2, .02, .9)$. For type (7), the phase shift $\phi \sim 2 \cdot \pi \cdot \mathcal{N}(0, \sigma^2)$ changes pulse wise, with modulation strength $\sigma \sim \mathcal{N}(.175, .05^2, .05, .3)$.

For type (8), two cyclic pulse trains with a fixed phase shift $\phi \sim 2 \cdot \pi \cdot \mathcal{N}(.5, .2^2, .2, .8)$ are added. The maximal amplitude of the shifted pulse train is $h \sim \mathcal{N}(.5, .2^2, .1, .9)$.

Listening experiment

Seven expert listeners participate in a 3 alternative forced choice (3AFC) discrimination experiment. Six of the listeners have a long experience in auditory clinical voice assessment, and the remaining one has a long experience in listening tests conducted in clinical audiology. The stimuli are presented via a loudspeaker Type 8020B (Genelec Oy, Iisalmi, Finland) at comfortable loudness. Three stimuli are presented in each trial. One of the three stimuli is randomly selected to be of a different type than the other two stimuli. The listeners are asked to auditorily identify the different one. The listeners are allowed to listen to all stimuli arbitrarily often. Playback is

controlled and answers are recorded by the listeners via a graphical user interface. Ten triplets are presented for each of the 28 pairs of GAW types, resulting in a total of 280 triplets. Correct and incorrect answers are counted, and correct rates in % are obtained.

RESULTS

The overall correct rate is 71.6 %. The listener with the highest correct rate achieved 80.4 %, whereas the one with the lowest achieved 58.9 %. Table 1 reports the correct rates with respect to GAW type pairs. Pairwise correct rates range from 37.1 % to 100 %. Correct rates are significantly higher than guessing rates in 21 out of 28 GAW type pairs.

Table 1: Correct rates (%) regarding the pairwise auditory discrimination of GAW types. Bold underlined rates are nonsignificantly better than guessing rates (χ^2-Test, 28-fold Bonferroni corrected, p=0.05 significance level).

		GAW type						
		2	3	4	5	6	7	8
GAW type	1	67.1	88.6	78.9	67.1	**51.4**	**55.7**	82.9
	2		84.3	**37.1**	**41.4**	72.9	78.6	74.3
	3			84.3	88.6	94.3	91.4	**37.1**
	4				**44.3**	71.4	70.0	80.0
	5					62.9	70.0	78.6
	6						**55.7**	100.0
	7							94.3

DISCUSSION AND CONCLUSION

The majority of the GAW types were auditorily distinguishable. However, types involving periodic modulations were not easily distinguished, and often confused with diplophonia (types 2, 4, and 5). Also, types that contain randomness are often confused with each other (types 1, 6, and 7). Finally, fixed phase differences are easily confused with bicyclicity (types 3 and 8). These results appear to be plausible for the following reasons: Regarding to confusion of periodic modulations with diplophonia, periodic modulations may produce a second pitch, which is the predominant perceptual cue for identifying diplophonia. Doubled pitches may be caused by periodic modulations, as well as by adding a second pulse train with a different fundamental frequency. Our experiments suggest that listeners are not able to distinguish these two causes auditorily. Also, listeners are not sensitive to the shape of the modulation functions in case modulation is periodic. Sawtooth-shape modulations may be auditorily perceived as brighter than sinus-shape modulations owing to a larger number of stronger partials for the former. However, this auditory property was not communicated to the listeners before the experiment. One may hypothesize that training may increase listeners'

sensitivity to the shapes of modulation functions in the future.

Regarding the confusion of signals with randomness, it was observed that these signals are often pooled by listeners into one category designated as noisy. One may hypothesize that extra pulsed voices with low extra pulse rates may be distinguished from noisy voices, because voices with low extra pulse rates sound crackling rather than noisy. A threshold extra pulse rate may exist, which separates crackling voices from raspy/noisy voices, because rare extra pulses are perceived as individual auditory events rather than continuous noise. Further, one may investigate in the future if modulation noise may be auditorily distinguishable from additive noise.

Regarding the confusion of bicyclic signals and signals with fixed phase differences, it was observed that a significant subset of waveforms obtained with the synthesizer for fixed phase differences is bicyclic. This overlap of stimuli sets explains the confusion. Also, it was observed that both of the signal classes sound rather normal, especially when compared with other GAW types, which sound rather abnormal. However, the auditory distinction between voicing types 3 and 8 from normophonic voice was not systematically tested in this experiment.

Modulated phase shifts of the left and the right vocal fold explain pulse shape and pulse height modulations, i.e., shimmer. This modelling option is an alternative perspective on shimmer, which may be interpreted as a result of the superposition of two cyclic waveforms, each having its own jitter. A next step of investigation may be to quantitatively comparing synthesized GAWs with observed natural ones.

REFERENCES

[1] P. Aichinger, I. Roesner, M. Leonhard, D. Denk-Linnert, W. Bigenzahn, and B. Schneider-Stickler, "A database of laryngeal high-speed videos with simultaneous high-quality audio recordings of pathological and non-pathological voices," in *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 2016, vol. 10, pp. 767–770.

[2] P. Aichinger, I. Roesner, J. Schoentgen, and F. Pernkopf, "Modelling of Random Extra Pulses During Quasi-Closed Glottal Cycle Phases," in *Models and Analysis of Vocal Emissions for Biomedical Applications*, 2017, vol. 10, pp. 129–133.

[3] P. Aichinger, "Diplophonic Voice - Definitions, models, and detection," PhD dissertation, Graz University of Technology, Austria, 2015.

VALIDATION OF A NONINVASIVE SYSTEM TO OBSERVE GLOTTAL OPENING AND CLOSING: EXTERNAL PHOTOGLOTTOGRAPH (ePGG)

Angélique Amelot[1], Darshan Sathiyanarayanan[1], Shinji Maeda[1], Kyioshi Honda[2], Lise Crevier-Buchman[1,3]

[1] Université Sorbonne-Nouvelle Paris 3 - Laboratoire de Phonétique et de Phonologie (LPP)-UMR7018, Paris, France
2 School of Computer Science and Technology, Tianjin, China
3 Hôpital Européen Georges Pompidou, Paris, France

Keywords: External PhotoGlottoGraph; ElectroGlottoGraphy; Video-endoscopy; Open Glottis

INTRODUCTION

Investigation of vocal fold movements required invasive instrumentations like Electromyography (EMG), Video-endoscopy or Transillumination. External lighting and sensing PGG (ePGG), recently developed at LPP [1] [2], relies on transillumination technique. It consists of illuminating the cavity above the glottis through the exterior neck skin surface with infrared LED (IR-LED) placed above the glottis and recording medially under the glottis variations of light intensity modulated by vibrations of or by abduction/adduction of the vocal folds with the 20 kHz sample rate. Compared with previous photo-glottographic techniques, this device has the definite advantage that both IR-LED and the photosensor are positioned on the speaker's exterior neck surface (*see Fig. 1*). Thus, the data can be noninvasively acquired.
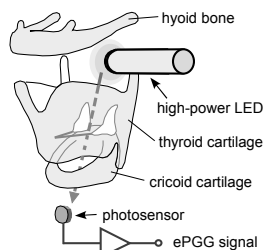
Fig. 1: External lighting and sensing photoglottography (ePGG) system: an IR-LED is placed on the skin on the neck, between hyoid bone and thyroid cartilage.

The ePGG system was recently tested on two mechanical larynx replicas [3]. Our current study is to validate and calibrate the outputs of ePGG on human speakers. We compared ePGG data with data from high-speed and standard video-endoscopies, and from electroglottography (EGG). The goal of the study is i) to validate ePGG signals by a multimode analysis and ii) to understand the relationships between the vocal folds movements and the degree of opening of the glottis indicated by ePGG and by EGG.

METHODS

Subjects
Two healthy subjects, one female (speaker 1) and one male (speaker 2) have participated in the data acquisition at the European Hospital of George Pompidou at Paris with the help of a medical doctor.

Instrumentations, corpus, and analysis
a) ePGG and EGG (EG2PCX2 Glottal Enterprises) signals, and high-speed video-endoscopy (Richard Wolf GmbH) images (4000 images/s) are simultaneously recorded as shown in Fig. 2. Audio signal is recorded by a microphone integrated in the video-endoscopy system. An additional external microphone is used. Its signal and those from ePGG and EGG are saved in a PC via a multichannel USB data card. These two audio signals allow us to make the time-alignment of the video-images sequence, and ePGG and EGG signals. For the comparison, frame-by-frame variations of glottal aperture are detected with the help of a MATLAB script resulting in a glottal aperture signal (GA). Because of the requirement of strong light for the high-speed operation, we are obliged to employ a rigid straight endoscope (*see Fig 2*), which limit the corpus to an isolated stationary high vowel [i]. The point of interest, therefore, becomes the comparison of glottal pulse shapes during the vowel [i]. By the way, in both this session and the next one b), the white LED light coming from the endoscope is used as the light-source for ePGG.
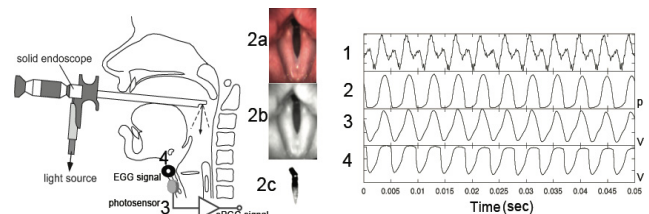
Fig. 2: Setup to acquire simultaneously (1) Acoustic signal in Volts, (2) images from high-speed video-endoscopy (4000 images/s): (2a) Original image, (2b) Filtered image, (2c) Detected GA and (2), GA, in pixels (3) ePGG signal in Volts, and (4) EGG in Volts.

b) In this session we focus our attention on slow abduction/adduction of vocal folds related to the production of vowel-consonant sequences. In such case, the high-speed video-endoscopy can be replaced by a standard video-endoscopy with the speed of 25 images/s. We use a flexible fiber-endoscopy, which is inserted through the nasal cavity. This, therefore, allows a speaker to utter any vowel-consonant sequence. The corpus in this

session consists of [VCV], [CVCV], and [CVC], where C=[s, z, p, b] and V=[a, i, u]. Each target word is repeated three times in the carrier sentence "Je dis ___ trois fois" ("I say ___ three times"). So, if a target word is [sas], then the utterance becomes "Je dis sas sas sas trois fois".

RESULTS

a) ePGG, High-Speed video-endoscopy and EGG
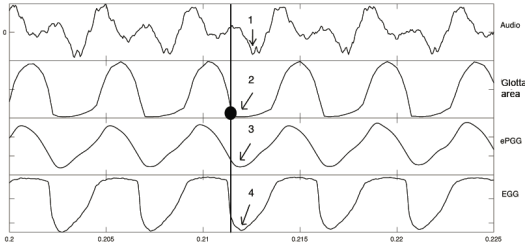
Fig. 3: Extract of the production of [i] for the speaker 1, From top to bottom: Acoustic signal, GA, ePGG, and EGG in which the polarity is inverted.

At the black dot (pointed by arrow 2) in Fig. 3, GA exhibits a strong discontinuity as if the closing vocal folds are suddenly blocked, creating a glottal closure. This occurs in every glottal cycles. The closure seems to be sustained for a while before a rapid opening of the glottis. According to the acoustic theory, the vocal tract must be excited at the closure onset by a negative pulse. Indeed, the audio signal indicates a sudden shoot (pointed by the arrow 1), which occurs about .5ms after the glottal closure onset. This .5ms delay must be the sound propagation time from the glottis to the microphone (about 20 cm), reinforcing the notion of the glottal closure onset and the closed glottal quotient.
However, we don't see those at all in ePGG and in EGG. In Fig. 3, both ePGG and EGG vary continuously exhibiting a turning point, at the arrow 3 and 4, respectively, from glottal closing to opening. We hypothesize, during the closure, that the tissues inside vocal fold are continuously deforming which is invisible to the endoscopic imaging but "visible" to ePGG and to EGG. The deformation of the tissues causes variations in the light transmission across closed glottis (ePGG sensitive) and that in the surface contact area of the vocal folds (EGG sensitive).
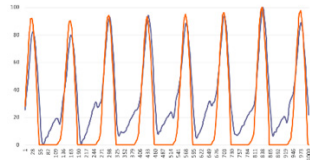
Fig. 4: Extract of normalizations of ePGG data and GA during the production of [i] for the speaker 2 (blue line is the ePGG data red line is GA from high-speed video-endoscopy (2000 images/sec).

The Fig. 4 shows how ePGG can be comparable to GA. In the ePGG curve we can recognize the closure onset as well as the closed glottis quotient, which are absent in the ePGG in Fig.3. The possible explanation is the gender difference.

Presumably on the one hand, the male subject has the thick and dense vocal-fold tissues and tight closure of the VF that tend to suppress the light going through the tissues. On the other hand, the female speaker has the thin and low dense vocal-fold tissues that leads to the result already described above.

b) We observe that ePGG can respond not only to the oscillation of the vocal-fold vibrations but also to abduction/adduction of the glottis for the consonant production. For example, during the production of an unvoiced fricative, ePGG can provide interpretable data, but EGG cannot as shown in Fig.5.
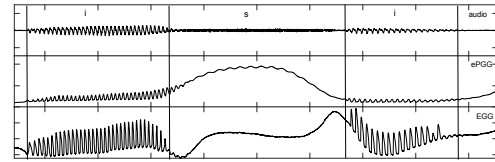
Fig. 5: From top to bottom, Audio signal [isi], ePGG, and EGG speaker

CONCLUDING REMARKS

We feel that ePGG is a versatile method for observing both glottal vibrations during voicing sounds and large glottal abduction/adduction movements in transitions between a consonant and a vowel. We could not identify glottal closure quotients on the ePGG. This appears to be because the light can go through thin vocal-fold tissues during the closure. EGG has shown a similar response in the same situation. In such a case, we need to resort to an invasive high-speed video-endoscope. However, simple and clean looking data obtained could be due to its blindness to the deformation of internal vocal-fold tissues behind the "closure".

ACKNOWLEDGMENTS

REFERENCES

[1] K. Honda and S. Maeda, "Glottal-opening and airflow pattern during production of voiceless fricatives: {A} new non-invasive instrumentation.," *J. Acoust. Soc. Am.*, vol. 123, no. 5, p. 3788, 2008.

[2] J. Vaissière, K. Honda, A. Amelot, S. Maeda, and L. Crevier-Buchman, "Multisensor platform for speech physiology research in a phonetics laboratory," *J. Phonetic Soc. Japan*, vol. 14, no. 2, pp. 65–78, Sep. 2010.

[3] A. Bouvet, A. Van Hirtum, X. Pelorson, S. Maeda, K. Honda, and A. Amelot, "Calibration of external lighting and sensing photoglottograph," Dec. 2017.

INVESTIGATION OF ACOSUTIC DIMENSION USE IN DIALECT PRODUCTION: MACHINE LEARNING OF SONORANT SOUNDS FOR MODELING ACOUSTIC CUES OF AFRICAN AMERICAN DIALECT

Meisam K. Arjmandi[1], Laura C. Dilley[1], Suzanne E. Wagner[2]

[1] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA
[2] Department of Linguistics, Michigan State University, East Lansing, MI, USA

INTRODUCTION

Listeners can rapidly draw inferences about the likely background of a speaker – including his or her dialect and racial background – within milliseconds of hearing a voice [1]–[3]. Purnell et al. [4] further showed that from just the word "*hello*", listeners not only draw inferences about the racial background of a talker, but this can lead to discrimination. The accuracy of this perceptual recognition of dialect was better than chance [4]. However, it is still not clear which acoustic cues and phonetic contexts facilitate this rapid inference about dialect and racial background.

Speech is the outcome of a dynamic interaction between vocal folds vibratory patterns and patterns of articulatory gestures and movements in the vocal tract. Dialect modulates the phonatory and articulatory patterns during speaking, leading to distinct cross-dialectal representations [5]. African American English (AAE) is a dialect spoken by many of the approximately 45 million African Americans. It was shown that formant dynamics are informative for separation of AAE from Standard American English (SAE) dialect [6], but the degree of contribution of other acoustic dimensions has not yet been investigated. This study aims to investigate the acoustic dimensions relevant to glottal source and/or the vocal tract that are most informative for separation of AAE from SAE dialect. The degree of informativity of these acoustic cues in AAE-vs-SAE classification across different phonological contexts are also addressed in this study.

METHODS

Materials

Six speakers (3 AAE and 3 SAE), all from Lansing, Michigan, participated in a sociolinguistic interview. Tokens of vowels conditioned on certain phonological contexts were identified and labeled by analysts in each audio recording. The chosen contexts were selected to control for coarticulation to ensure reasonable comparison across two dialects. The selected contexts involved a closed syllable with a sonorant coda consisting of liquids or nasal of /l/, /r/, /n/, or /m/ or an open syllable (*i.e.*, no consonant coda). Target stretches of speech consisted of vowel (V) or vowel-consonant (VC) sequences. The total duration of

analyzed speech was 183.3 sec. (100.1 sec. AAE and 83.4 sec. SAE). Sonorant sounds (*e.g.*, V and VC) were a focus, because these carry substantial acoustic cues relevant to dialect identification [5].

Acoustic Measures

Four general categories of acoustic features were calculated to characterize acoustic variations in multiple dimensions with respect to their informativeness in AAE vs. SAE dialect separation: *(1) Speech-based features:* These measures reflect the behavior of both glottal source and vocal tract. These features were *H1-H2, H1-A1, H1-A1, H1-A2, A1-A3, H1-A3,* which were calculated by amplitudes of the 1st and 2nd harmonics (*cf. H1* and *H2*) relative to each other and to the amplitude of the 1st, 2nd, and 3rd formants (*cf., A1, A2, A3*), and *spectral slope* which reflects the rate of decline in spectral amplitude. *(2) Vocal tract features (F1, F2, and F3):* These features represent the natural resonances of the vocal cavity. *(3) Voice quality (VQ) measures (jitter, shimmer, mean F_0, STD F_0, fraction of unvoiced frame (FOUF), mean harmonic-to-noise ratio (HNR)):* These measures characterize the quality of voice during V and VC pronunciation. (4) *Duration and energy* (RMS)*:* These measures were used to characterize energy and linguistic stress in the voice [7]. A schematic of the implemented approach to extracting the acoustic features within these four categories is shown in Fig. 1.
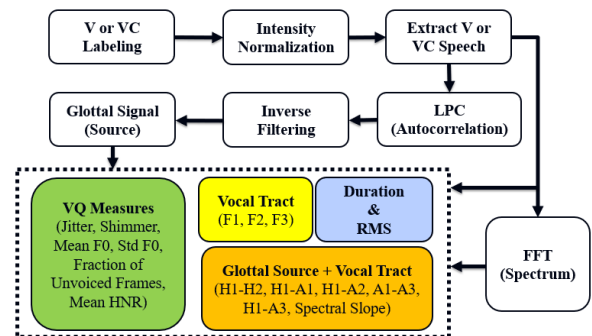
Figure 1. Schematic of the implemented method to extract the acoustic features of the four categories from V and VC speech sounds.

Analysis

The informativeness of these acoustic cues was evaluated at two steps of a multi-dimensional space feature evaluation and a machine learning-based approach. We

first evaluated the informativeness of these acoustic features individually to identify their discriminative strength in AAE-vs-SAE dialect separation. *Mahalanobis distance* [8] was used as a non-probabilistic measure to rank the acoustic features. This metric evaluates the distance of a feature in a multi-dimensional space from the mean of the class. We finally trained a support vector machine (SVM) on a 17-dimension feature space to identify how much the acoustic dimensions formed by these features are informative in AAE-vs-SAE dialect separation.

RESULTS AND DISCUSSION

Fig. 2 shows the average contribution of acoustic features across seventeen phonological contexts based on the ranking given by the *Mahalanobis distance* measure. The figure clearly demonstrates the important role of *vocal tract features* (*F1*, *F2*, and *F3*) in the separation of AAE from SAE dialect [9]. This figure also suggests that the main contributions come from *speech-based features* and *vocal tract features*. The results of testing the trained SVM on unseen test samples in Fig. 3 shows that sonorant sounds of /aɪ/, /ɪl/, and /il/ provide the most informative acoustic cues for the SVM classifier to distinguish AAE from SAE.
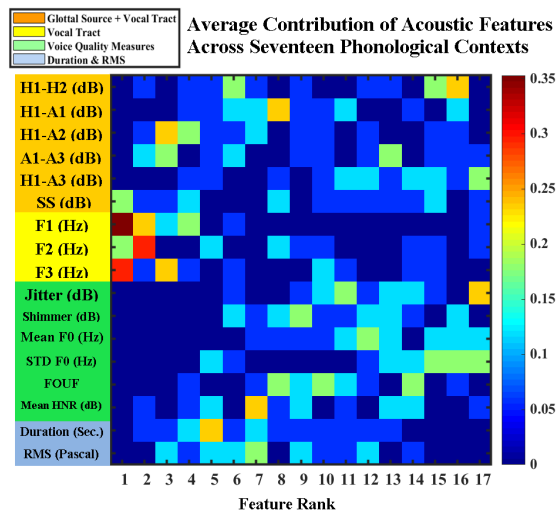
Figure 2. Average contribution of acoustic features in different categories in AAE-vs-SAE separation.
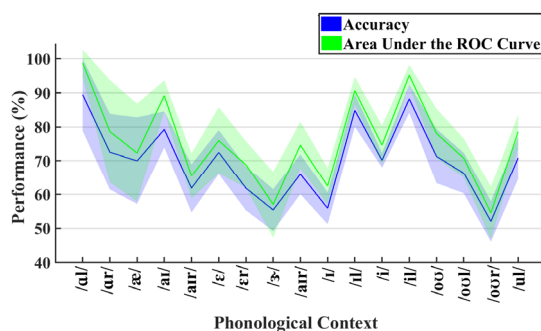
Figure 3. Accuracy and AUC of SVM in classification of SAE and AAE dialect.

CONCLUSION

The results from this study suggest that rapid recognition of AAE dialect as distinct from SAE dialect is facilitated through the interaction of acoustic features representing both phonatory behaviors and articulatory gestures. Among various acoustic cues, formants in V and VC contexts provide substantial acoustic information for recognition of AAE from SAE, supporting our prior findings [6]. Investigating the acoustic cues from continuous speech, including obstruents, rather than merely sonorant regions, can be planned for future studies.

REFERENCES

[1]   B. Munson, "The acoustic correlates of perceived masculinity, percieved femininity, and percieved sexual orientation," *Lang. Speech*, vol. 50, no. 1, pp. 125–142, 2007.

[2]   S. Lattner and A. D. Friederici, "Talker's voice and gender stereotype in human auditory sentence processing - Evidence from event-related brain potentials," *Neurosci. Lett.*, vol. 339, no. 3, pp. 191–194, 2003.

[3]   M. Scharinger, P. J. Monahan, and W. J. Idsardi, "You had me at 'Hello': Rapid extraction of dialect information from spoken words," *Neuroimage*, vol. 56, no. 4, pp. 2329–2338, 2011.

[4]   T. Purnell, W. Idsardi, and J. Baugh, "Perceptual and Phonetic Experiments on American English Dialect Identification," *J. Lang. Soc. Psychol.*, vol. 18, no. 10, pp. 10–30, 1999.

[5]   R. A. Fox and E. Jacewicz, "Cross-dialectal variation in formant dynamics of American English vowels," *J. Acoust. Soc. Am.*, vol. 126, no. 5, pp. 2603–2618, 2009.

[6]   M. K. Arjmandi, L. Dilley, and Z. Ireland, "Applying pattern recognition to formant trajectories: A useful tool for understanding African American English (AAE) dialect variation," *J. Acoust. Soc. Am.*, vol. 141, no. 5, pp. 3980–3980, May 2017.

[7]   D. B. Fry, "Duration and Intensity as Physical Correlates of Linguistic Stress," *J. Acoust. Soc. Am.*, vol. 27, no. 4, pp. 765–768, 1955.

[8]   S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2008.

[9]   P. Ladefoged and D. E. Broadbent, "Information Conveyed by Vowels," *J. Acoust. Soc. Am.*, vol. 29, no. 1, pp. 98–104, 1957.

VOCAL REHABILITATION – LESSONS LEARNED FROM A 30-YEAR ANALYSIS OF PUBLICATIONS IN THE JOURNAL OF VOICE

Mara Behlau

Centro de Estudos da Voz (CEV), Universidade Federal de São Paulo – UNIFESP and Insper, São Paulo, Brazil

Keywords: **Voice; Vocal quality; Basic Sciences; Research – mbehlau@cevbr.com.br**

INTRODUCTION

The voice area is essentially multiprofessional and multidisplinary. Even if this poses major challenges in terms of knowledge advancing, on the other hand it offers a unique and composed scenario with inputs from many sources. The goal of this presentation is to share an analysis considering The Journal of Voice - JoV as a source of information. For this presentation, I decided to focus in 3 aspects: 1. The Challenge of the G.P. Moore Lecture given in 2016; 2. The Evolution on Vocal Rehabilitation Papers at the JoV and 3. The Basic Science Awards at JoV.

1. THE CHALLENGE OF THE G.P. MOORE LECTURE GIVEN IN 2016

50 articles were compiled to honor the contributions made by authors to the Journal of Voice during a period of thirty years, from 1987 to 2016 [1]. These articles were categorized in 5 topics: 1. Normalcy of the larynx and voice; 2. The clinical speech-language pathologist (SLP) evaluation; 3. The patient's perspective; 4. The core of vocal rehabilitation; and 5. Behavioral versus Organic Dysphonias. Normalcy of larynx is a complex aspect since the physical appearance of a healthy larynx varies across individuals with normal voices; similarly, normal voice is difficult to define because it is not a binary descriptor, but itis a tendency, a measure of a multifaceted perceptual phenomenon with cultural influence and highly dependent of the context for its analysis. The clinical SLP, voice specialist, was greatly benefited by computerized technology and the average clinician can count on a voice laboratory to register the patients' voice and monitor the outcome of treatment, a situation that was before limited to only few research and university centers. However, one of the major paradigm shifts in this area was not a consequence of sophisticated measures but on the inclusion of the patients' opinion about the impact of a voice problem on his/her quality of life, considering a mix of factors such as disadvantage, reduced participation and the perception of limitation in daily, social and professional activities besides difficulties in the expression of the emotions. Vocal rehabilitation is a non-linear process that involves a combination of direct and indirect approaches. Direct approaches produced a higher level of scientific evidence that indirect approaches. Clinicians are aware that the type of approach is only one aspect that must be considered when planning a vocal therapy, other issues may play an important role, such as the personality of both the patient and the clinician, coping style, adherence of treatment and readiness for changes and self-control, of voice usage. Finally, voice clinicians face many challenges: to understand the variability of cases and to develop scientific and cultural competence are surely central regarding the clinical attitude. Two different etiologies of voice problems need to be didactically considered since they require different treatment approaches: behavioral and organic dysphonias. The role of the vocal behavior as the origin of the voice problem is what separates one category from the other. Vocal behavior can be defined [1] as the set of vocal reactions in response to interpersonal relationships in the environment in which the individual lives. Changing vocal habits and behaviors are important focuses for the treatment of this category of voice problems. Although organic dysphonias, such as vocal fold paralysis and hypophonia due to Parkinson's disease are not necessarily the result of harmful vocal behaviors, both can benefit from vocal rehabilitation and a better vocal output can be obtained.

2. THE EVOLUTION ON VOCAL REHABILITATION PAPERS AT THE JoV

A review of the complete JoV collection from the CEV library was performed and concluded with a list of 144 manuscripts dealing with vocal rehabilitation[2]). These articles were classified into 4 general groups: theoretical articles (N=10), review articles (N=12), outcome studies with dysphonic patients (N=75, with 7 *c*ase report studies, 35 studies evaluating immediate effects or duration of voice exercises and 9 studies with defined methods/programs*)* and effects of vocal rehabilitation and related factors (N=47). The distribution throughout the years reveals that the last decade has received the highest number of publications (90 articles from 2007 to 2017; 29 from 1997 to 2006 and 23 from 1987 to 1996) with the highest concentration, 26 contributions, in 2017. The JoV is a real international platform that counted with the participation of authors from 29 countries to disseminate their result of researches and clinical observations. The highest number of contributions came from the USA (56 articles), followed by Brazil (15 articles), The Netherlands (7 articles), Australia (6 articles), Belgium, Italy and Sweden (5 articles each). Fifteen publications were

credited to authors from more than one country, with the United States co-author in 6 publications. This recent clear growth of number of publications and international authorship is probably the result of many factors, however, the improvement in basic science and the dissemination of studies in the internet have helped the SLP working in the voice area to show the value of services provided via well designed studies. 49 different therapeutic approaches were explored in many publications. Mutiparametric and multidimensional effects of vocal rehabilitation evidenced that not a single measure can reflect the therapy outcomes. The analysis of the whole JoV collection regarding vocal rehabilitation revealed a clear development of studies design, with an improvement on the methodological description. Many philosophical orientations, from a symptomatic to a holistic approach, are present all over the decades, expressing the multiplicity of approaches all over the worlds, with some cultural preferences. Multiple approaches, specific programs and customized treatments, regular and intensively administered to many types of vocal problems produced a variable level of scientific evidence of positive results. Effect size and a more robust statistical analysis must follow in the years to come.

3. THE BASIC SCIENCE AWARDS AT THE JoV

The JoV best papers awards are given annually, since 2007 and the selection is made by the editorial board members. The Basic Sciences category has received 12 recognitions (there was a tie result in 2017; see references 3 to 14). 7 out of the 12 papers deal with wound healing aspects of the vocal folds, both on acute stage of injury as well as long-term consequence of a scar tissue, which can create a devastating voice problem. A better understanding of regenerative models and tissue reorganization can produce changes in prevention, rehabilitation (behavioral agents) and surgery of the scarred vocal fold (direct manipulation of tissue or injection of materials). Other topics, such as subglottal pressure or vocal efficiency ratio (2 articles), high-speed imaging algorithm or protocol to measure vocal fold movements (2 articles) and the morphology of vocal tract (1 article) were also awarded. Authors from 7 countries took part in these articles and the USA has contributed with most of publications (8 studies, 2 in conjunction with authors from other countries). Clear clinical implications were mentioned in all studies, which marks the profile of this Basic Science category at JoV.

Fostering interactions among voice scientists, clinicians, and surgeons should be focus of international conferences in the voice area, such as the 11th ICVPB. I received this invitation as a reflection of the understanding of the organizers on their responsibility to enhance exchange of knowledge and experience in the field. I feel honored for this opportunity to share some thoughts in the voice area.

REFERENCES

[1] Behlau M. The 2016 G. Paul Moore Lecture: lessons in voice rehabilitation? Journal of Voice and clinical practice. J Voice 2018; J Voice. 2018 [Epub ahead of print]

[2] Behlau M, Carroll L. 2018. Vocal rehabilitation or voice therapy at Journal of Voice: a 30-year analysis of publications. J Voice. 2018 /accepted/

[3] Branski R, Perera P, Verdolini K, Rosen C, Hebda P, Agarwal S. Dynamic Biomechanical Strain Inhibits IL-1β–induced Inflammation in Vocal Fold Fibroblasts. J Voice 2007; 21:651–660.

[4] Graupp M, Kiesler K, Friedrich G, Ainodhofer H, Gruber H-J, Kieslinger P, Saxena A, Hirano S. Gugatschka M. Vocal Fold Fibroblast Response to Growth Factor Treatment is Age Dependent: Results From an In Vitro Study. J Voice, 2014; 28:420-423.

[5] Hirano S, Minamiguchi S, Yamashita M, Ohno T, Kanemaru S, Kitamura M. Histologic Characterization of Human Scarred Vocal Folds. J Voice, 2009; 23:399–407.

[6] Liu Y, Xi H, Xing W, Gu J. Aquaporin Changes in Compound 48/80 Induced Inflammatory Sublaryngeal Edema in Rat. J Voice 2012; 26:815.e17-815.e23.

[7] Mainka A, Platzek I, Mattheus W, Flishcer M, Müller A-S, Mürbe D. Three-dimensional Vocal Tract Morphology Based on Multiple Magnetic Resonance Images Is Highly Reproducible During Sustained Phonation. J Voice, 2017; 31:504.e11-504.e20.

[8] Moon I, Park K, Kim H, Lee S. Utility and Safety of Commercially Available Injection Laryngoplasty Materials in a Rabbit Model. J Voice, 2015; 29:125-128.

[9] Poburka B, Patel R, Blss D. Voice-Vibratory Assessment With Laryngeal Imaging (VALI) Form: Reliability of Rating Stroboscopy and High-speed Videoendoscopy. J Voice, 2017; 31:513.e1 - 513.e14.

[10] Sivasankar M, Fisher K. Vocal Folds Detect Ionic Perturbations on the Luminal Surface: An In Vitro Investigation. J Voice, 2008; 22:408-419.

[11] Sundberg J, Scherer R, Hess M, Muller F, Granqvist S. Subglottal Pressure Oscillations Accompanying Phonation. J Voice, 2013; 27:411-421.

[12] Thibeault S, Klemuk S, Chen X, Quinchia Johnson B. . In Vivo Engineering of the Vocal Fold ECM With Injectable HA Hydrogels—Late Effects on Tissue Repair and Biomechanics in a Rabbit Model. J Voice, 2011; 25:249-253.

[13] Titze I, Maxfield L, Palaparthi A. An Oral Pressure Conversion Ratio as a Predictor of Vocal Efficiency. J Voice, 2016;30:398-406.

[14] Zhang Y, Bieging E, Tsui H, Jiang J. Efficient and Effective Extraction of Vocal Fold Vibratory Patterns from High-Speed Digital Imaging. J Voice, 2010; 24: 21-29.

REGARDING PHONATORY CONDITIONS IN THE IN VIVO CANINE LARYNX USING AUTOMATED, GRADED STIMULATION

David A. Berry, Dinesh K. Chhetri

Department of Head & Neck Surgery, University of California, Los Angeles, California, USA

Keywords: Phonation; Neuromuscular stimulation; In Vivo Canine Model

Abstract:

Objectives: Phonation threshold pressure (P_{th}) is often used as an objective measure of the relative ease of phonation, and fundamental frequency (F_0) variations carry important linguistic information in speech. Many previous studies of *in vivo* canine phonation have reported values of P_{th} which are much higher than values typically reported for human subjects, and similarly, many previous studies of in vivo canine phonation have reported much more restricted ranges of F_0 than are typically reported for human subjects. Such studies are often cited to argue that human and canine phonation possess significant functional differences. The hypothesis of this research is that using methods of graded stimulation to the laryngeal nerves and/or muscles, the ranges of P_{th} and F_0 in the *in vivo* canine laryngeal model are more comparable to that of human phonation that previously reported.

Methods: In this study, we investigate P_{th} and F_0 as a function neuromuscular stimulation to the laryngeal nerves and/or muscles of the in vivo canine larynx using an automated technique for graded stimulation, as described previously (Chhetri, Neubauer & Berry, 2012).

Results: In the study, across over a dozen experiments, P_{th} was observed in the range of 0.2 to 2.8 kPa and F_0 was observed in the range of 85 – 850 Hz.

Conclusions: The hypothesis was confirmed that the ranges of P_{th} and F_0 in the *in vivo* canine larynx in these experiments are much closer to ranges typically reported for human subjects, In particular, P_{th} was observed to be as low as 0.2 - 0.3 kPa (which closely matches the lower range of P_{th} traditionally reported for human subjects), and F_0 was observed to span over 3 octaves. Previous studies of in vivo canine phonation may have been limited by an inadvertent, exclusive focus on hyper-stimulation of the TA (thyroarytenoid) muscle.

LOMBARD EFFECT IN A RESTAURANT SETTING

Pasquale Bottalico, Kelly Kost, Emilie Palacios, Alison Perlman, Angelica Wozniak

Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, Champaign, Illinois, USA

Keywords: Lombard effect; Restaurant settings; Noise levels; Vocal effort

INTRODUCTION

According to the 2016 Zagat State of American Dining report, 25% of restaurant customers consider noise the most irritating component of dining out. Restaurants are not merely eating sites, they also serve as business and social hubs. Typical noise levels in restaurants range from 60 dBA to 85 dBA. In a noisy environment, speakers unconsciously attempt to maintain a level of speech that allows them to be understood (Lombard effect). Lazarus [1] stated that the speech level rises with noise level with a slope of 0.3-0.6 dB per noise level rise of 1 dB for all disturbing noise exceeding 40-50 dB(A). However, up to a noise level of 30-40 dB(A), noise level has minimal effect on the speech level [2]. Tang *et al.* [3] estimated the Lombard effect by measuring the variation of noise level with the number of occupants in a university staff canteen. From their results, they deduced that occupants began to raise their voices when the background noise level exceeded 69 dB(A). Hodgson *et al.* [4] proposed a model for predicting speech and noise levels, and the acoustical conditions for verbal communication including the Lombard effect, in eating establishments. The Lombard slope resulting from the proposed model was 0.69 dB/dB.

The aim of this study is to identify the specific point when the restaurant noise causes vocal discomfort for customers. The study also seeks to discover how willing a customer would be to spend their time and money in a restaurant, depending on the varying intensities of noise levels in the surrounding environment.

METHODS

Human Data

In this study, the effects of typical restaurant noise varying between 35 and 85 dB(A) on (1) vocal effort (quantified as SPL), (2) the amount of self-reported disturbance in the communication due to noise and vocal control, and (3) the willingness to spend time and money for a meal, were evaluated. The speech of the 14 male and 14 female talkers (18-28 years; mean 21 years) was recorded in a sound-attenuated booth with typical restaurant noise. In order to simulate a real communication setting, participants were seated in the booth facing a human listener, positioned at a 1 m distance. The 10 noise conditions were presented in a random order. Participants were asked to read the text (Rainbow passage) to the listener using the following instructions: "Each time, I [the listener] would like you to pretend that we are talking in a restaurant and you are telling the story to me. Make sure that I understand you equally well each time." The listener was present in the booth during the entirety of the experiment. After each reading of the text, participants were asked to answer the three questions on a visual analog scale about the experience of talking in the various noise level conditions.

Instrumentation and Measurement

Speech was recorded by a head-mounted microphone (Shure Beta 54 WBH54) in eleven noise conditions between 35 and 85 dB(A) in 5 dB increments. The noise levels for the 11 conditions were measured by placing the microphone in the position of the participants' ears. Typical restaurant noise was emitted by two directional speakers placed at 45°, 1.5 m in front of the participant. The gain of the playback software of the studio monitors was modified in order to obtain increments of 5 dB in the position of the participants' ears. The reverberation time (T20) at mid-frequencies in the room was 0.05 s, while the background noise was 22.5 dB(A).

Analysis

MATLAB (R2017a) was used for speech signal analysis. In each condition, the equivalent SPL was measured. For each condition, the mean value of the SPL was obtained per subject. For each subject, the average of SPL among the conditions was computed and subtracted from each mean SPL values for that subject (termed ΔSPL).

Statistical analysis was conducted using R version 3.1.2. Four piecewise linear (also called segmented or broken-line) models were fit to the response variables SPL, self-reported discomfort and willingness to spend time and money with the predictor, noise level, using the segmented package in R.

RESULTS

The ΔSPL was measured at each of the 11 noise levels between 35 and 85 dB(A), as shown in Figure 1. A piecewise linear model was fit to the response variable, ΔSPL, and the predictor, Ln. The slope of the lower segment was 0.31, and the upper was 0.54, with a change-point identified in Ln at 57.3 dB(A) (CI 95% lower: 53.0, CI % upper: 61.6) with an R-squared of 0.91.

Self-reported communication disturbance was measured at each of the 11 noise levels (Figure 1). A piecewise linear model was fit to the response variable, disturbance (%) and

the predictor, Ln. The slope of the lower segment was 0.74, and the upper was 2.51, with a change-point in Ln identified at 52.2 dB(A) (CI 95% lower: 48.0, CI % upper: 56.4) and an R-squared of 0.80.

The willingness to spend time in the restaurant was measured at each of the 11 noise levels, as shown in Figure 1. A piecewise linear model was fit to the response variable, discomfort (%) and the predictor, Ln. The slope of the lower segment was -0.65, and the upper was -2.24, with a change-point in Ln identified at 51.3 dB(A) (CI 95% lower: 46.9, CI % upper: 55.7) with an R-squared of 0.77.

The willingness to spend money in the restaurant was measured at each of the 11 noise levels, as shown in Figure 2. A piecewise linear model was fit to the response variable, discomfort (%) and the predictor, Ln. The slope of the lower segment was -0.63, and the upper was -1.72, with a change-point in Ln identified at 52.5 dB(A) (CI % lower: 44.7, CI % upper: 60.2) with an R-squared of 0.60.
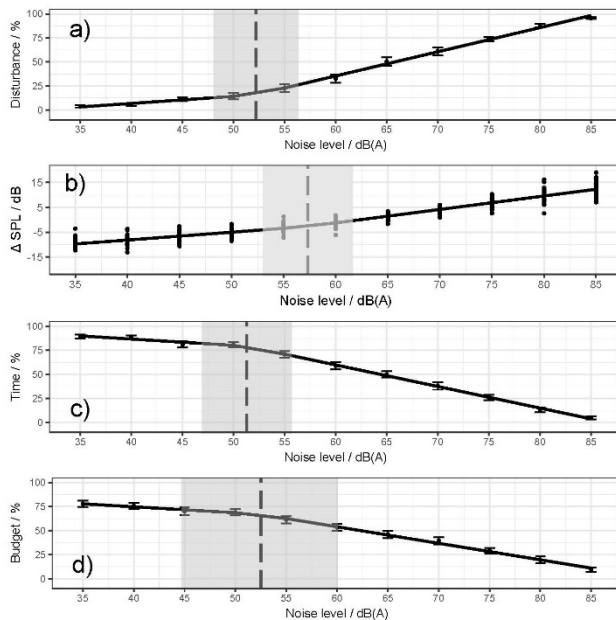
Fig. 1: Relationship between the level of the noise in dB(A) and self-reported level of communication disturbance (a), voice level (b), the willingness to spend time (c) and money (d), where the error bands indicate the standard error. Vertical dashed lines mark the change-points.

DISCUSSION

Based on the reported results, which are comparable to previous work, it can be claimed that vocal level (effort) and disturbance increase as background noise increases, while the willingness to spend time and money in a restaurant decreases as background noise increases.

The hypothesis of a starting point for the Lombard Effect was not verified. However, as background noise increased, change-points could be identified in the slope of the increase in vocal level (Lombard Effect), disturbance,

and willingness to spend time and money in a restaurant. Regarding the objective measure of vocal effort, a change-point of the Lombard effect was identified at a noise level equal to 57.3 dB(A), while Tang *et al.* [3] deduced that occupants began to raise their voices when the background noise level exceeded 69 dB(A).

The difference in the changing point level could due to different factors: in the present study the participant and the research assistant were the only people in the room, this configuration probably inhibited possible automatic mechanisms in speech regulation related to privacy. This could be an explanation for the early changing point of the Lombard effect. Moreover, the results of Tang *et al.* [3] are based on prediction models and the authors pointed that their model is limited to the fact that the raised voice threshold was chosen empirically and thus it may be specific to the present experimental conditions. After the changing-point, the Lombard slope found in the present study was 0.54 dB/dB. This compares well with the Lombard slopes of 0.2 to 1 dB/dB reported in the literature, and with the Lombard slope of 0.69 dB/dB reported by Hodgson *et al.* [4] based on prediction model.

Regarding the subjective measures, the change-point for disturbance was lower at a noise level equal to 52.2 dB(A). Participants started to become (more) disturbed by noise at a lower noise level than with the associated change-point of the Lombard effect. Similar noise levels (51.3 dB(A) and 52.5 dB(A)) are also triggering a decrease in the willingness to spend time and money in a restaurant.

CONCLUSION

To improve the acoustic environment of a restaurants, background noise levels should be lower than 50-55 dB(A). This will minimize the vocal effort of patrons and the disturbance in their communication. Concurrently, this will increase business for the restaurant since patrons would be willing to spend more time and money to eat in a restaurant with a background noise lower than 50-55 dB(A).

REFERENCES

[1]. H. Lazarus, "Prediction of Verbal Communication is Noise—A review: Part 1," Applied Acoustics, 19(6), 439-464 (1986).

[2]. P. Bottalico, I. I. Passione, S. Graetzer, and E. J. Hunter, "Evaluation of the starting point of the Lombard effect," Acta Acustica United with Acustica 103(1), 169-172 (2017).

[3]. S. Tang, D. W. Chan, and K. Chan, "Prediction of sound-pressure level in an occupied enclosure," The Journal of the Acoustical Society of America 101(5), 2990-2993 (1997).

[4]. M. Hodgson, G. Steininger, and Z. Razavi, "Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments," The Journal of the Acoustical Society of America 121(4), 2023-2033 (2007).

STEADINESS OF SINGING SCALES BY UNTRAINED ADULT FEMALES

Michelle Bretl[1], Ronald Scherer[1]

[1] Department of Communication Sciences and Disorders, Bowling Green State University, Bowling Green, Ohio, U.S.A.

***Keywords*: Aerodynamics; Electroglottography; Register Transitions; Untrained Singers**

INTRODUCTION

As a singer, register transitions can be challenging to navigate. Not only does a singer need to perceive where the transition is occurring, but also apply any technique or adjustment in order to smooth the transition. Within register transitions, one may experience various types of vocal instabilities. Identified gross vocal instabilities include jumps to vocal fry and aphonia episodes, , and the natural instability of the adolescent voice change. Parametric measures are used to define instability in the voice, such as abrupt pitch change, an intensity jump, and variations in the electroglottographic (EGG) signal.

The primary aim of this study was to identify voice instabilities during singing by untrained female participants and describe the singing relative to changes in f_o, airflow, intensity, adduction, and acoustic spectra. It was expected that smooth scales would show little variation in signals and little indication of register transitions, while unsteady scales would be revealed by variations in objective signals and measures. Additionally, it was anticipated that subtle characteristics of unsteadiness might be more obvious within the objective analyses rather than auditory perception.

METHODS

Participants and Tasks

Five untrained young adult female singers completed the tasks in the study. Participants were chosen based on four criteria: no history of private voice lessons (choral experience acceptable), the ability to produce an octave scale in a lower and an upper range, relatively normal speech and voice, and not a smoker. Participants had experiences in choirs, bands, and other music groups.

Singing "naturally", the participant began with a lower scale that crossed the lower *passaggio* (register transition). Then the singer moved to a higher scale that crossed the upper *passaggio*. On each scale, the participant sang on a continuous /a/ vowel on one breath and did the same on the /i/ vowel. Additionally, the participants sang each vowel at two different dynamic levels, *mezzo piano* (medium soft) and *mezzo forte* (medium loud).

Instrumentation and Measurement

A microphone system was used to record the audio signal for acoustic analysis. The aerodynamic system with a circumferentially vented flow mask (Glottal Enterprises, model MSIF-2) was used to record broadband flow [1].

The EGG signal (Kay Elemetrics) was analyzed using custom software (Sigplot, a MATLAB-based program) to infer vocal fold contact measures and disruptions in phonation.

Analysis

The microphone signal provided measures of fundamental frequency and intensity, analyzed through Praat software. Additionally, the first time derivative of the fundamental frequency was used. Narrowband spectra were used to compare spectral slope from the first formant peak to the third formant peak over the scales.

The smoothed airflow signal was used to visualize changes in the airflow. The EGG signal was examined for its waveform and changes in the waveform shape. It was also smoothed to examine indications of vertical (DC) shifts in the signal (and thus vertical shifts of the larynx).

A total of 13 different instabilities were realized throughout all of the analyzed scales. Instabilities were judged based on what was heard in the audio signal, and what was seen in the airflow and EGG signals. Instabilities were also identified in the plots of the first derivative of the fundamental frequency. Several instabilities can be identified in the plot below.
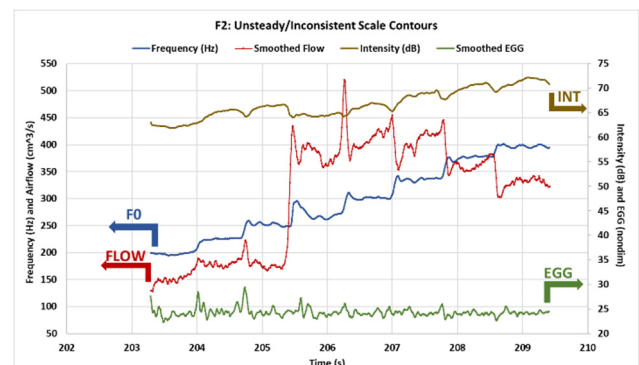
Fig. 1: Unsteady Scale Contours. Intensity (dB), Fundamental Frequency (Hz), Smoothed EGG, and Smoothed Airflow (cm³/s). Sung on /a/ vowel, loud intensity level, from G3 to G4 pitches.

RESULTS

The recorded scales were categorized based on perceptual smoothness of the auditory signal. Both investigators agreed upon an appropriate level for each scale based on perceptual quality and perceptual dynamics of the scale. Scales were classified into three groups,

marked as Smooth, Middle, and Unsteady. The groupings were nearly equal: 36.4% (32 of 88 scales) were perceived in the Smooth category, 27.2% (24 of 88 scales) were perceived to be in the Middle category, and 36.4% (32 of 88 scales) in the Unsteady category. These three overarching groups allowed for determination of salient features of unsteadiness based on observations and analyses of the acoustic, airflow, intensity, and EGG patterns.

In the Smooth group, the salient features tended to be the following: consistent voice quality throughout the scale, minimal pitch and intensity variations (with the exception of vibrato), and consistent airflow patterns and strategies. However, in the Smooth group, there were often some instabilities that went undetected auditorily, including pitch overshoots and large airflow variations.

In the Middle group, the salient features tended to be the following: perceptible but fairly smooth registration shifts, minor pitch bobbles, non-vibrato pitch variations, and quality changes across the scale. Scales perceived to be part of the Middle group generally had overwhelmingly smooth features with one or two instances of perceptible changes or instabilities.

In the Unsteady group, the salient features tended to be the following: abrupt voice quality changes often attributed to registration shifts, large pitch overshoots and variations (unrelated to vibrato), aphonic segments, loss of the EGG signal waveform due to presumed laryngeal height or tissue contact change, and unexpected intensity changes.

Correspondence among the various instabilities is seen in **Table 1**. The instabilities listed in the table include all the different instabilities found in the signals from all participants' tokens. The number of instabilities within a single scale ranged from zero to fourteen. For the scales classified in the Smooth category, the majority of scales had zero to three instabilities within the scale. For the scales classified in the Middle category, the majority had four to six instabilities within the scale. For the scales classified in the Unsteady category, the majority had six or more instabilities within the scale. The scales perceived to be smooth had fewer measurable instabilities than the scales perceived to be unsteady.

DISCUSSION AND CONCLUSION

The results allowed prioritization of measures relative to importance or salience. Some measurable instabilities included: abrupt frequency shifts or f_o variations (other than vibrato) with continuous sound; aphonic segments with an absence of sound; glottal adductory shifts relative to EGGW measures of the EGG signal; unusual DC shifts of the EGG signal; abrupt intensity changes; and inconsistencies of f_o rate of change. Airflow instabilities included abrupt flow changes due to abrupt adductory changes and/or hypothesized transglottal pressure changes.

Table 1: Instabilities across all analyzed scales.

Group Categorization (Smooth [S], Middle [M], Unsteady [U])	S	M	U
Total scales	32	24	32
Frequency: Aphonic segment	0	0	14
Intensity: change by >10dB at once	0	0	29
Airflow: change >150 cm^3/s across pitch change	5	9	12
Frequency: abrupt shift or wobble (unrelated to vibrato)	3	8	8
Perceptual voice quality change	1	11	18
Airflow: Inconsistent airflow changes on pitch changes	8	8	16
Rate of Change: Pitch overshoot >375 Hz/s on pitch change	6	18	27
Rate of Change: Maximum positive peak to maximum negative peak difference >1600 Hz/s	7	11	17
Rate of Change: Max positive to min positive peak OR max negative to min negative peak difference >900 Hz/s	4	6	12
EGG: amplitude change by >30%	19	14	12
EGG: EGGW50 value change by >0.1	13	26	15
EGG: Inconsistent direction or pattern of DC shifts on each pitch change	14	17	23
EGG: waveform signal dropout	3	9	17
Total Instabilities	83	137	220

The primary findings of this project are the following:

1. Untrained singers can sing smoothly across octave scales including passaggio transitions – 36.4% of the total number of scales analyzed were perceived as smoothly sung.

2. Rate of change of f_o reveals unsteadiness that may be related to subtle aspects of register change and vocal control.

3. Smoothed (average) airflow and EGG signals also appear to be sensitive measures for register change and vocal control, especially relative to adduction.

4. The primary unsteadiness variable was an aphonic segment and corresponding abrupt and large intensity reduction.

5. Register shifts from modal to head register appeared to correspond to steeper spectral slopes, increased smoothed airflow, and decreased EGG waveform height and width.

6. Some of the objective measures appear to be more visually and numerically salient than their corresponding auditory perceptual presence relative to unsteadiness in the production of scales, especially when more obvious perceptual unsteady characteristics are not present.

REFERENCES

[1] Rothenberg, M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *J Acoust Soc Am.* 1988;84:511-529.

FROM VOCAL EFFORT TO VOCAL FATIGUE. WHAT DOES THE LITERATURE SAY?

Lady Catherine Cantor-Cutiva, Russell Banks, Mark Berardi, Brittany Johnson, Rebecca Clawson, Salma Martínez, Eric Hunter

Department of Communication Sciences and Disorders, Michigan State University, East Lansing, Michigan, United States

***Keywords*: Vocal fatigue, vocal effort**

INTRODUCTION

Although vocal fatigue is a term that has been frequently used in the literature, there is a lack of consensus on how vocal fatigue can or should be assessed and defined [1]. Indeed, the use of the term *vocal fatigue* (as well as the related terms, *vocal effort*, *vocal load, and vocal loading*) has varied depending on the voice scientists and the study. Previous studies have reported how vocal fatigue is induced by "vocal loading" tasks, which require vocal effort from the talkers [2]. This close relationship between vocal fatigue, effort, and loading may make it difficult to establish a clear definition. For example, any of these aspects (vocal fatigue, effort, loading) have been explained, *in general terms*, as an increased muscular tension, with variations in voice acoustic parameters, and related with contextual variation, such as room acoustics [3, 4].

Moreover, since vocal fatigue, effort, load, and loading are multidimensional concepts, there is a need for establishing "holistic" definitions for these concepts which can be accepted by the community of interest. However, before such definitions can be established, it is first necessary to identify how and where these terms are used. This requires a specialized research method to obtain reliable indicators related to quantity of the published papers on these topics in order to define which dimensions should be included in these "holistic" definitions; such a method is called bibliometric studies. To the best of the authors knowledge, up no bibliometric study concerning vocal fatigue, vocal effort, vocal load, and vocal loading is available to date. Therefore, the aim of the current study was to provide an updated and comprehensive description of the "intellectual structure" and the evolution of research on vocal fatigue, vocal effort, vocal load, and vocal loading. The results presented in this paper add useful information to the field of voice, not only because it is the first study to apply bibliometric techniques to our field, but also because, in so doing, it complements and expands the findings of previous studies that have approached the subject from a qualitative perspective, such as reviews of literature [3, 5]. Also, our results may contribute to the second step in the process of evidence-based practice (i.e., finding the evidence) as advised by the American Speech-Language-Hearing Association (ASHA, 2004) and would lay the groundwork necessary to revisit the definitions of the terms.

METHODS

Objectives

1. To analyze the bibliometric indicators of scientific publications on vocal fatigue and vocal effort published in databases available in the Library system of Michigan State University (Intellectual structure);

2. To build a comprehensive definition of vocal fatigue and vocal effort from the review of literature on these topics (evidence-based definition).

Methods

This study was designed based on an exploratory and retrospective analysis conducted by way of a bibliometric study. We used 4 databases to identify articles on vocal fatigue and vocal effort. The databases consulted were: PUBMED, SCIELO, EMBASE, and COCHRANE. We included articles written in English, Spanish and Portuguese. Then, a systematic review was conducted and information regarding the definitions and assessment methods of vocal fatigue, and vocal effort were extracted from each paper. As a second step, a latent semantic analysis was performed to map keywords into a "concept" spaces (fatigue and effort). Next, a comparison in these spaces were performed in order to find the smallest set of keywords that spans all the reviewed articles per topic. A comprehensive definition was defined for each topic including the most recurrent keywords found in the reviewed manuscripts.

RESULTS

Intellectual structure

Figure 1 shows the evolution of publications per topic. Before 1990, 100% of the publications reported results on research about vocal effort. This tendency changed during the 90´s when research in vocal fatigue, vocal load, and vocal loading started to be published. From 2000 to 2017, vocal fatigue has been the leading topic among the four, with about 45% of the published manuscripts on this topic.

In total, 113 manuscripts on vocal fatigue and vocal effort included definitions of these topics. Around 56% (n=63) were on vocal fatigue, whereas 44% (n=50) were on vocal effort. In total, 33 words were identified as keywords for the definition of vocal fatigue, whereas 42 were identified for the definition of vocal effort.

Fig. 1: Evolution of publications per year and topic.

As shown in the figure 2, room acoustics, speech adjustments, and distance to the interlocutor are the three most common keywords related to the "concept" space of vocal effort.

Fig. 2: Keyword analysis for concept space of "vocal effort".

On the other hand, vocal hyperfunction, voice disorders, vocal effort, and vocal loading were identified as part of the "concept" space of vocal fatigue (Figure 3).

Fig. 3: Keyword analysis for concept space of "vocal fatigue".

Evidence-based definition of vocal effort

Vocal effort is a multidimensional concept that has been defined from 5 different perspectives. From the perceptual and physiological point of view, vocal effort is commonly defined as an increase in vocal loudness and strain in voicing that may be linked with an increased subglottal (tracheal) pressure, and a higher cervical muscle tension. From the point of view of the person with vocal effort, some of the most common symptoms are pain/discomfort when speaking, feeling of tightness in the throat, and/or sounding tearful. From an acoustic perspective, it has been suggested that vocal effort is associated with an increase in fundamental frequency and its variation, the first formant, and sound pressure levels and its variation in voice. From the psychological perspective, vocal effort was associated with mood changes and therefore considered mood induced, also individuals with vocal effort were identified as less confident. A more holistic approach to vocal effort including the environmental aspects consider three main aspects: distance to the interlocutor, background noise, and time in vocal use.

Evidence-based definition of vocal fatigue

Vocal fatigue was commonly defined as a vocal symptom or deficit, the result of prolonged voice use or vocal loading and improve with vocal rest. One subset defined vocal fatigue as a perceptual condition: Vocal fatigue was defined as increased vocal effort, reduced control of voice flexibility and quality, and/or a weak voice. The other subset defined vocal fatigue as a physiological condition: Vocal fatigue was the result of laryngeal muscle and tissue fatigue resulting in laryngeal discomfort and reduced range and control of fundamental frequency and intensity.

CONCLUSION

Definitions of key aspects, such as vocal fatigue or vocal effort are not always clarified in published research in this field. This lack of a uniform definition has resulted in the use of similar terminology to define different conditions. Therefore, it is extremely important to use a standardized terminology in the field of voice in order to facilitate the comparison of results.

REFERENCES

1. Nanjundeswaran, C., et al., *Vocal Fatigue Index (VFI): Development and Validation.* J Voice, 2015. **29**(4): p. 433-40.
2. Boucher, V.J. and T. Ayad, *Physiological attributes of vocal fatigue and their acoustic effects: a synthesis of findings for a criterion-based prevention of acquired voice disorders.* J Voice, 2010. **24**(3): p. 324-36.
3. Welham, N.V. and M.A. Maclagan, *Vocal fatigue: current knowledge and future directions.* J Voice, 2003. **17**(1): p. 21-30.
4. Whitling, S., *Vocal loading and recovery.* 2016, Doctoral thesis, Lund University, Department of Logopedics, Phoniatrics and Audiology, Lund, Sweden.
5. Fujiki, R.B. and M.P. Sivasankar, *A Review of Vocal Loading Tasks in the Voice Literature.* J Voice, 2017. **31**(3): p. 388 e33-388 e39.

THE RELATION BETWEEN VOCAL FRY AND ROOM ACOUSTICS AMONG BILINGUAL ENGLISH-SPANISH SPEAKERS.

Lady Catherine Cantor-Cutiva,[1], Pasquale Bottalico[2], Eric J. Hunter[1]

[1] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA
[2] Department of Speech and Hearing Science, University of Illinois Urbana-Champaign, Champaign, Illinois, USA

Abstract:

Purpose: Define the changes on occurrence of vocal fry among bilingual English-Spanish speakers under different virtual simulated acoustic conditions.

Methods: Thirty-four bilingual English-Spanish young adults participated in this cross-sectional study. Participants produced two speech samples (one in English and one in Spanish) under nine different room acoustic conditions (simulated). The text in English consisted on the first six sentences of "The Rainbow Passage", equal to about 30 seconds of speaking. The text in Spanish was a fragment of the text "El Caballero de la Armadura Oxidada" (The Knight in Rusty Armor, a standardized text in Spanish) with similar duration of speaking.

Result: Native English speakers tend to produce vocal fry more often under the "No Noise" condition compared with noisy conditions. On the contrary, production of vocal fry among native Spanish speakers seem to be no affected by background noise conditions.

Conclusions: Bilinguals English-Spanish speakers tended to produce less vocal fry when they were speaking in Spanish with babble background noise than when they were speaking in English with "No background noise".

References

Cantor-Cutiva LC, Bottalico P, Hunter E. Factors associated with vocal fry among college students. *Logopedics Phoniatrics Vocology*. August 2017:1-7. doi:10.1080/14015439.2017.1362468

Cantor-Cutiva LC, Bottalico P, Ishi CT, Hunter EJ. Vocal Fry and Vowel Height in Simulated Room Acoustics. *FPL*. 2017;69(3):118-124. doi:10.1159/000481282

LOMBARD EFFECT AND MUSCLE TENSION DYSPHONIA: ACOUSTIC, AERODYNAMIC, BIOMECHANICAL AND CORTICAL CHANGES

Christian Castro[1,2], Pavel Prado [2] Daphne Marfull[2], Alba Testart[3], Alejandro Weinstein[3], Lucia Zepeda[3], Victor M. Espinoza[2,4], Matías Zañartu[2]

[1] Department of Speech, Language and Hearing Sciences, Universidad de Valparaíso, Valparaíso, Chile
[2] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile
[3] Department of Biomedical Engineering, Universidad de Valparaíso, Valparaíso, Chile
[3] Department of Speech Pathology, Universidad de Playa Ancha de Ciencias de la Educación, Valparaíso, Chile
[4] Department of Sound, Universidad de Chile, Santiago, Chile

Keywords: e.g. Lombard Effect; High-Speed Video endoscopy, Electroencephalography; Auditory Feedback

INTRODUCTION

One of the most common yet least understood hyperfunctional voice disorders is muscle tension dysphonia (MTD). MTD is defined as poorly regulated activity of the intrinsic and extrinsic laryngeal muscles leading to high levels of stiffness and tension in the vocal folds with an excessive laryngeal and paralaryngeal contractions in absence of known structural or neurologic disorders [1]. The etiology of MTD remains unclear, but most patients report overly increased vocal efforts when speaking in noisy environments, in which the Lombard Effect (LE) is elicited. The LE is an unconscious response when talking in noisy environments [2] that has been well documented in healthy subjects. Previous studies report that subjects in noisy environment increase vocal loudness, pitch and articulatory aspects due to the disruptions in the auditory feedback [2]. However, the dynamics of biomechanical and aerodynamic parameters as well as the cortical mechanisms involved in the control of the voice production during the LE have not been fully described. The aim of this study is to explore if acoustic, aerodynamic, biomechanical, and neurophysiology parameters describing the Lombard effect in MTD patients differ from those obtained healthy controls.

METHODS

Ten healthy volunteers and MTD patients with normal audiometric assessments were included in this study. Subjects were asked to utter a series of vowels and syllables that were displayed on a screen. Images were presented for 3 seconds with a 3-second pause between stimuli. A total of 40 stimuli were distributed in 8 series of 5 images. Vocal function was assessed by high-speed videoendoscopy (HSV), aerodynamic measurements, and acoustic measurements in three conditions: Basal, Lombard (in noise) and recovery (5 minutes after the end of LE conditions). LE was elicited with speech noise, presented at 80 dB HL. The acoustic signal was obtained using a condenser microphone (B&K 4961). Aerodynamic signals were recorded using a circumferentially vented mask (Glottal Inc, MS-110) and HSV was recorded using

high-speed camera (Photron, SA-X2) connected to a rigid endoscope (KayPentax, 9106).

Glottal airflow measures included peak-to-peak amplitude of the unsteady airflow (AC Flow), maximum flow declination rate (MFDR) and open quotient (OQ), all computed using custom MATLAB codes. Other parameters included mean voice intensity (SPL), posterior glottal closure (PGC), and asymmetry phase between vocal folds (AP). Two-way ANOVAs ($p < 0.05$) were performed to analyze the dynamics of the different parameters, using the group and the acoustic condition as factors.

The number and duration of the utterances were slightly modified in the experimental protocol for conducting electroencephalography (EEG). The electrical activity was recorded from 64 scalp-electrodes (10/20 system), using a BioSemi ActiveTwo 64-channel EEG acquisition system. The amplitude of the N1-P2 complex of the event related potentials (ERP) was compared among conditions, also using a two-way ANOVA ($p < 0.05$). Cortical activations were estimated using standardized low-resolution brain electromagnetic tomography (sLORETA).

RESULTS

Healthy and MTD subjects responded to speaking in noise elicited the LE. Both groups significantly increased the intensity of their voice in comparison with the intensities obtained in the basal condition. Remarkably, the greatest increase in intensity was obtained in the MTD group. After 5-minutes in quiet, the vocal intensity of healthy volunteers was similar than that obtained in the basal condition. The vocal intensity of MTD patients also decreased in the recovery condition as compared with the levels obtained in noise. However, the intensity of their vocalizations, 5 minutes after the end of LE conditions, was significantly higher in comparison with the basal condition, as clearly seen in Figure 1. Regarding aerodynamic measures, both control and MTD groups revealed significantly higher AC Flow and MFDR for the Lombard condition in comparison with the basal condition. Nevertheless, when comparing the basal and recovery conditions, aerodynamic parameters of the healthy volunteers returned to their basal level, 5 minutes after the

end of LE conditions. Although the MFDR of patients significantly decreased after returning to a quiet environment, they did not returned to those obtained previous the induction of the LE.

Figure 1: Acoustic and aerodynamic measures for control and MTD group in Basal, Lombard and Recovery conditions. Dotted lines represent the beginning and the end of the Lombard condition, respectively
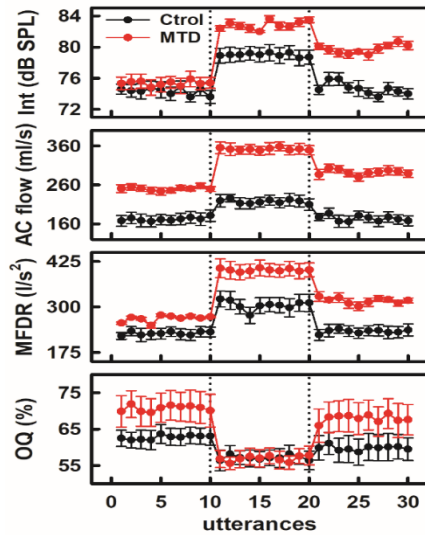
The ERPs of a representative subject of the control group are displayed in Figure 2. The cortical response was characterized by the increased N1-P2 complex amplitude of the ERP elicited by the subject own vocalizations in noise, in comparison with that obtained in the basal condition. The response amplitude decreased in the recovery condition, but did not returned to its basal level.

Figure 3 shows that the auditory feedback of the subject's own voice induced the activation of left temporal and frontal areas, including the Broca area, Wernicke area, primary auditory cortex, and primary motor cortex (44, 22, 41/42, and 4 Brodmann areas, respectively). Increased activations were obtained during the LE as compared with those obtained in the basal condition. Five minutes after returning to quiet, the activation of language areas located in the temporal decreased. However, the activity of the parietal areas associated to language increased (Brodmann areas 39 and 40).

Figure 2: ERP of one healthy subject in Basal, Lombard and Recovery conditions.
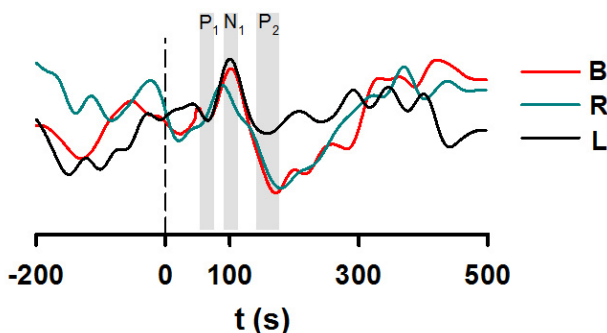
Figure 3: Cortical activation of one healthy subject in Basal (B), Lombard (L) and Recovery (R) conditions. Contrasts are represented in the lower panel. Increased and decreased cortical activations in the reference condition are represented in red and blue, respectively
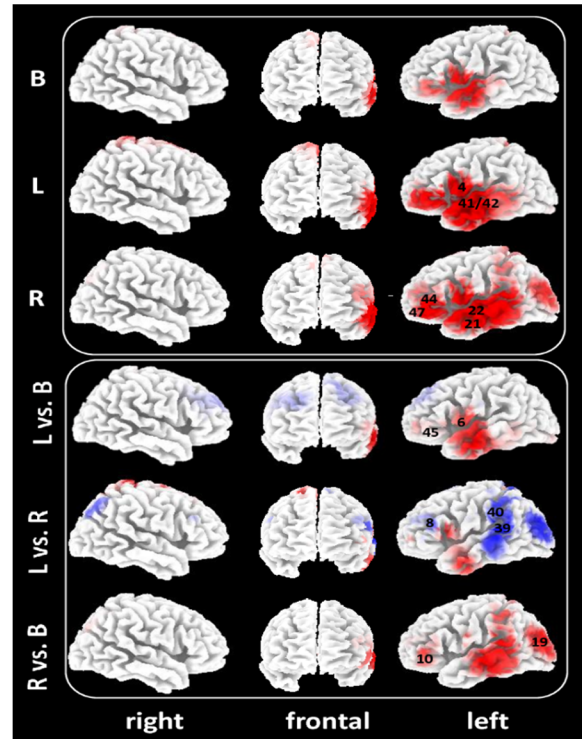
DISCUSSION AND CONCLUSIONS

Preliminary results suggest that subjects with MTD might be more sensitive to the Lombard effect, since they generate a greater variation in all biomechanical measures associated to hyperfunction phonatory patterns. Moreover, the MTD subjects have more difficulties to return to baseline conditions once the noise is removed in comparison with control subjects. These results are consisting with previous studies on pitch shift reflex, in which different adaptive responses between control and MTD subjects are reported [3]. Thus, these results suggest that MTD patients have a disrupted auditory-motor control integration during speech production.

REFERENCES

[1] Aronson, A. E. (1990). Clinical voice disorders: An interdisciplinary approach (3rd ed.). New York, NY: Thieme, Inc

[2] Garnier M, Henrich N. Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise. Computer Speech and Language 28 (2014) 580–597

[3] Stepp. C, Lester-Smith R, Abur. D, Daliri. A, Noordzij. P, Lupiani AA. 2017 Evidence or auditory-motor impairment in individuals with hyperfunction voice disorders.J Speech Lang Hear Res. 2017 Jun 10;60(6):1545-1550.

HUMAN VOCAL FOLD EPITHELIAL CELLS AND THEIR IMMORTALIZATION

Xia Chen and Susan Thibeault

Department of Surgery, University of Wisconsin-Madison, Madison, Wisconsin, United States

Keywords: **Human Vocal Fold Epithelial Cells; Cell Line; Telomerase Reverse Transcriptase; Transfection**

OBJECTIVE

Primary normal human vocal fold epithelial cell (hVFE) cultures are useful for studying vocal fold epithelial biology as well as epithelial mesenchymal interactions. Because of the limited ability of primary hVFE to proliferate *in vitro*, it is necessary to immortalize and establish cell lines that are suitable for long-term culture and large-scale *in vitro* experimentation. The purpose of this study was to isolate hVFE from human vocal fold tissue, and establish stable proliferative hVFE cell lines.

METHODS

Primary normal hVFE were isolated from healthy donors. Cells were steadily transfected by a retroviral vector containing human telomerase reverse transcriptase (hTERT) gene. We have characterized the morphological features and cell markers of E67T4 immunocytochemistry staining, analyzed hTERT gene expression and proliferative ability by PCR and MTT assays.

RESULTS

The E67T4 cell line showed typical "cobblestone" morphology of epithelial cells similar with primary hVFEs as Figure 1. Short-term proliferation rates of EE67T4 were similar with primary VFEs, but they maintained stable proliferative ability for eight months, whereas primary hVFE stopped proliferating after four passages. hTERT gene and protein expression were confirmed in E67T4 cell nucleus (Figure 2), and E67T4 showed same cell markers including cytokeratin 4, 13 and 14 with primary hVFE (Figure 3). DNA fingerprinting confirmed E67T4 was consistent with the presence of a single cell line. Karyotype analysis demonstrated E67T4 with some abnormal chromosome changes.

CONCLUSIONS

An immortalized hVFE cell line, E67T4 was established and characterized demonstrating consistent cell markers as primary hVFE, and constant proliferation rate for more than eight months comparing to primary hVFE stopping proliferation after four-time passing. This epithelial cell line can be used widely in biological applications in the larynx. We are presently optimizing 3D culture of these cells.

ACKNOWLEDGMENTS

Figure 1. The morphology features of immortalized and primary human vocal fold epithelial cells

Figure 3. Immortalized and primary VFE shared similar human VFE markers

Figure 2. The immortalized VFE was confirmed by ICC of cytokeratin 14 and human Tert
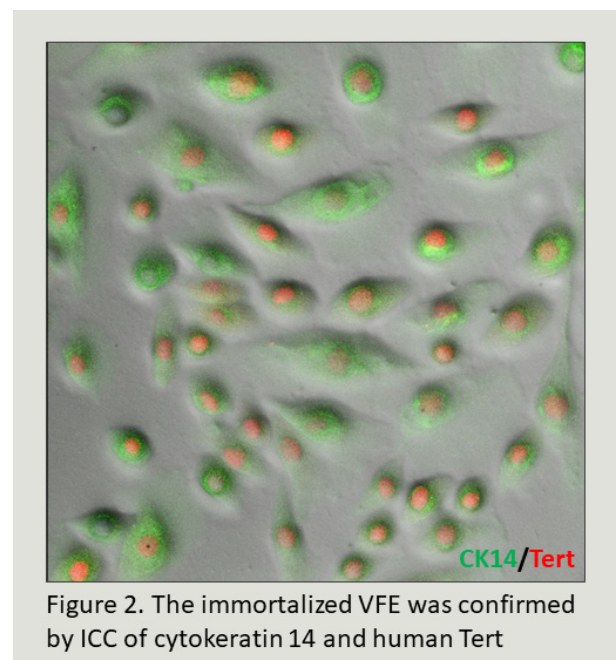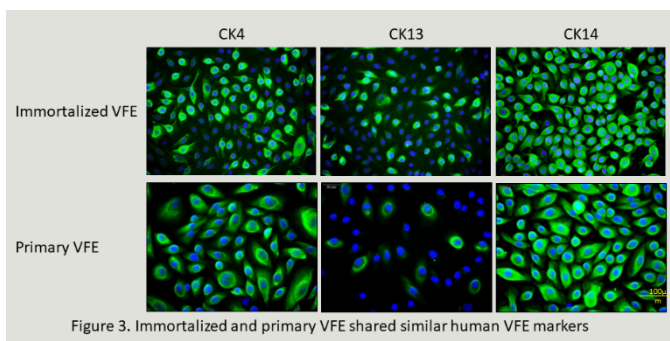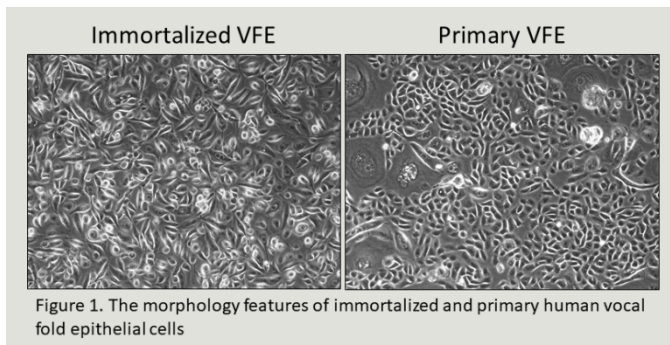
QUANTITATIVE COMPARISONS OF CORONAL PLANE GEOMETRY IN PHONATING HUMAN VOCAL FOLDS IN VIVO USING VCSEL OPTICAL COHERENCE TOMOGRAPHY

Lily Y. Chen[1], Giriraj K. Sharma[1,2], Christopher Badger[1,4], Ellen Hong[1], Li-dek Chou[1], Swathi Rangarajan[5], Theodore H. Chang[5], Christian H. Barnes[2], William B. Armstrong[2], Sunil P. Verma[2], Ram Ramalingam[5], Zhongping Chen[1,3], Brian J-F. Wong[1,2,3]

[1] Beckman Laser Institute, University of California, Irvine, Irvine, CA, USA
[2] Department of Otolaryngology-Head and Neck Surgery, University of California, Irvine, Irvine, CA, USA
[3] Department of Biomedical Engineering, University of California, Irvine, Irvine, CA, USA
[4] School of Medicine, University of California, Irvine, Irvine, CA, USA
[5] OCT Medical Imaging Inc., University of California, Irvine, Irvine, CA, USA

Keywords: Vocal Folds; Optical Coherence Tomography; Image Segmentation; Motion Analysis

INTRODUCTION

Clinical assessment of vocal folds (VF) conventionally involves laryngoscopy and/or videostroboscopy [1], both of which limit analysis to the surface topography. Though they are extremely well-suited for capturing lateral vocal fold motion, fail to capture vertical displacement. Comprehensive characterization of vocal folds in vivo necessitates a minimally invasive, cross-sectional imaging technology such as optical coherence tomography (OCT). OCT yields massive high-resolution data sets which provide a means of observing the behavior of various vocal fold tissue layers during phonation. Using OCT data, we have extensively analyzed and quantified subsurface structures and VF motion. Use of a VCSEL device for laryngeal OCT is essential to traverse the long path length from endoscope terminus to fold margin.

METHODS

Image Acquisition and Pre-Processing

Imaging sessions were completed in a clinical office setting. A 45° gradient index lens rod was connected to a swept-source based OCT system utilizing a 1310 nm center wavelength 200 kHz vertical cavity surface-emitting laser. Images were captured at a sampling rate of 800 a-lines per frame and 200 frames per second.

Raw interferometer spectral data was processed using custom acquisition software [2] to yield Doppler shift images and grayscale images for image visualization. Data series were reviewed identify data sets that featured full field vocal fold visualization. Using ImageJ, the scans were cropped to a general ROI containing the true vocal folds in order to reduce memory requirements and processing load. Additional image stabilization was completed using the OpenCV Template Matching algorithm as implemented in [3] to minimize bulk motion.

Analysis

Subsequent analysis was completed in Python. The grayscale images were smoothed using a total variation denoising Split-Bregman method [4-5]. The denoising method regularization parameter was selected as 0.1 (λ=0.1). Image intensity histograms of the denoised images were generated. Using the a priori knowledge that any remaining noise exhibits a normal distribution of low pixel intensity values, epithelium a medium intensity value distribution, and perceivable lamina propria (LP) the highest intensity value distribution, demarcating intensity values to separate the layers were identified. These values were utilized in the marching squares algorithm, as implemented in scikit-image [5], to contour the tissue layers.

With the contours of the epithelium and LP in sequential frames of motion, extensive geometric measurements could be made from every A-line in each scan (Fig. 1). Due to loss of signal in certain images because of bulk motion away from the focal point, some images did not have successful contours. To accommodate for this, the mean and standard deviation values for the layers were calculated after excluding measurements three standard deviations away from the initial mean value.

Vertical displacement was separately calculated as the average maximum displacement of the most medial top point of the through each frame. Compiling results from individual frames together yielded plots of VF thickness variations in space-time. In addition, a form of velocity of the vocal folds was tracked. 30 points were distributed equally on the top side of each epithelial contour in each frame. Displacement between the sets of points of consecutive frames was found and allowed for the calculation of velocity. Vocal fold motion could also be summarized through visualizing the contours in 3D using Matplotlib Python plotting library. Doppler images were assessed for banding patterns through color filtration to track the propagation of the mucosal wave.
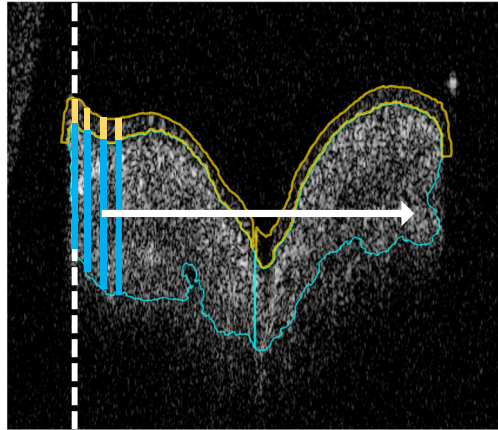
Fig. 1: Epithelial and LP layer thicknesses were measured by calculating the length of contour segments in each vertical A-line, as demonstrated respectively by the vertical gold and blue lines. Image is depicted with a 10.75:1 width-to-height ratio.

RESULTS

The results of geometric measurements have been summarized in Table 1.

Table 1: Summary of geometric measurements.

	Median (mm)	Mean ± St. Dev (mm)
L Epithelium	0.16	0.17 ± 0.07
R Epithelium	0.16	0.18 ± 0.08
Left LP	0.89	0.87 ± 0.29
Right LP	0.94	0.88 ± 0.26
L Vertical Disp.	1.10	1.06 ± 0.50
R Vertical Disp.	1.09	1.08 ± 0.57
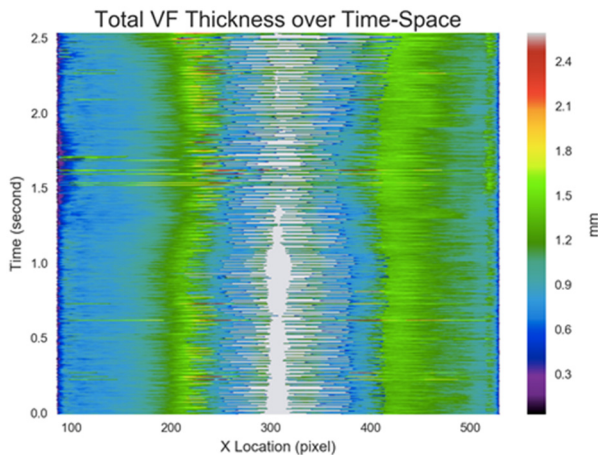
Fig. 2: A plot of the thicknesses of the LP and epithelium summed from every A-line in the same position of every frame is shown.

Surface displacement and vertical thickness of the vocal folds at various locations and times can be observed in plots such as Fig. 2. From plots of vocal fold velocity in time, we observed velocities over 1000 mm/s. Doppler analysis show axial spread of the mucosal wave penetrating approximately 3 mm into the VF.

DISCUSSION AND CONCLUSION

The findings for normal human vocal folds are consistent with existing reported measurements [6-8], although our perceivable lamina propria thickness measurements are less than histological reports. This may be attributed to the effects of light scattering, optical strength, and focus of view. With newly introduced space-time charts, comparisons can be made with abnormal vocal fold motion. It is expected that because abnormal phonation stems from structural changes in the vocal folds, the thickness and velocity space-time charts will highlight regions of interest. Furthermore, propagation of the mucosal wave is expected to deviate as well. With higher sampling frequencies, the resolution of these assessments will steadily improve. In tandem with OCT technology, our method can assist otolaryngologists in rapidly assessing patients' VF.

REFERENCES

[1] Mehta DD, Hillman RE. The evolution of methods for imaging vocal fold phonatory function. *Perspect Speech Sci Orofac Disord* 2012;22(1):5-13.

[2] Coughlan C, et al. In vivo cross-sectional imaging of the phonating larynx using long-range Doppler optical coherence tomography. *Sci Rep* 2016;6:22792.

[3] Tseng Q. Template Matching and Slice Alignment---ImageJ Plugins 2015.

[4] Wang X, et al. A two-step iteration mechanism for speckle reduction in optical coherence tomography. *Biomed Signal Process Control* 2018;43:86-95.

[5] Walt Svd, et al. scikit-image: Image processing in Python. *PeerJ* 2014;2:e453.

[6] Kaiser ML, et al. Laryngeal epithelial thickness: a comparison between optical coherence tomography and histology. *Clin Otolaryngol* 2009;34(5):460-466.

[7] Prades JM, et al. Lamina propria of the human vocal fold: histomorphometric study of collagen fibers. *Surg Radiol Anat* 2010;32(4):377-382.

[8] George NA, et al. Depth-kymography: high-speed calibrated 3D imaging of human vocal fold vibration dynamics. *Phys Med Biol* 2008;53(10):2667-2675.

AMNIOTIC FLUID FOR TREATMENT OF VOCAL FOLD INJURY IN AN IN VIVO RABBIT MODEL

Michael B Christensen[1,2], Jenny L Pierce[3], Phil Lee[4], Kristine M Tanner[5], Marshall E Smith[4]

[1] Department of Bioengineering, University of Utah, Salt Lake City, UT, USA
[2] National Center for Voice and Speech, University of Utah, Salt Lake City, UT, USA
[3] Department of Communication Sciences and Disorders, University of Utah, Salt Lake City, UT, USA
[4] Division of Otolaryngology/Head and Neck Surgery, University of Utah, Salt Lake City, UT, USA
[5] Department of Communication Disorders, Brigham Young University, Provo, UT, USA

Keywords: Injury; Hyaluronic Acid; Amniotic Fluid

INTRODUCTION

Hyaluronic acid-based (HA) injectables have been commonly studied as treatment materials in animal vocal fold injury models [1-7]. However, despite their common use, HA treatments are not able to fully restore tissue properties. Amniotic fluid (AF) has demonstrated healing potential in regenerative medicine in a wide variety of tissues, including the airway [8]. The current study aimed to investigate the treatment potential of injectable AF for vocal fold wound healing in an animal model and compared outcomes to an HA injectable.

METHODS

Materials

Two injectable materials were used in this study. HA injectable material was HyStem®-C Hydrogel (Ascendance Biotechnology, Inc., Alameda, CA). The AF injectable material was decellularized human AF developed by the University of Utah Department of Cell Therapy & Regenerative Medicine.

Injury and Treatment Groups

Three groups of six (N=18) male New Zealand white rabbits were included in the study. Procedures performed under general anesthesia delineated the three groups. Group 1 received vocal fold punch biopsy and immediate treatment of 0.15 mL AF injectable into the wound bed. Group 2 received punch biopsy and immediate treatment of 0.15 mL HA injectable. Group 3 received punch biopsy only, and served as the untreated control. The left vocal fold of each animal received biopsy only with Decker Rongeur 2x6mm cup forceps, or biopsy plus injection, and the right vocal fold of each animal served as its own uninjured control. Rabbit sacrifice and laryngeal dissection took place after six weeks to allow for wound healing and scar initiation.

Larynges were harvested immediately following necropsy and transferred to vials containing PBS for immediate transport for further dissection. Vocal folds were then harvested from the larynges. Removal of cartilage was ensured to avoid adding extra mass during rheology. Each dissected vocal fold was stored in PBS in a separate cryo vial, which was immediately flash frozen in liquid nitrogen. Cryo vials were then stored in a freezer at -80 °C until rheological analyses.

Rheology

Rheological analyses were performed on each of the 36 vocal folds. Following laryngeal dissection, vocal folds were stored at -80 °C until testing. Prior to testing, vocal folds were thawed and hydrated in a 1% phosphate buffered saline (PBS) solution at room temperature. Rheological testing was performed using a Bohlin Gemini-150 controlled stress rotational rheometer (Malvern Instruments Ltd, Worcestershire, UK) within an 8 mm parallel plate system. The temperature at the bottom plate was maintained at 37 °C using a water jacket attached to a F25-ME external heating and circulation unit (Julabo USA Inc, Allentown, PA). 220 grit sandpaper (Norton Abrasives, Worcester, MA) was attached to both plates with Permabond 105 (Permabond LLC, Pottstown, PA) to promote tissue adhesion to the plates. Sandpaper was replaced every three samples to prevent wear. Samples were placed between the plates and the upper plate was lowered in increments of approximately 100 μm until the normal force applied to the tissue remained greater than 25 g for at least 5 seconds. The gap size between plates was then decreased by an additional 20%. 1% PBS was applied to the tissue to ensure continued hydration. A controlled–stress program was used to complete frequency sweeps from 0.01-10 Hz. Ten sample points were recorded every decade. Viscoelastic properties of the tissue were determined by computing the elastic shear moduli (G') and viscous moduli (G") as a function of oscillatory frequency.

RESULTS
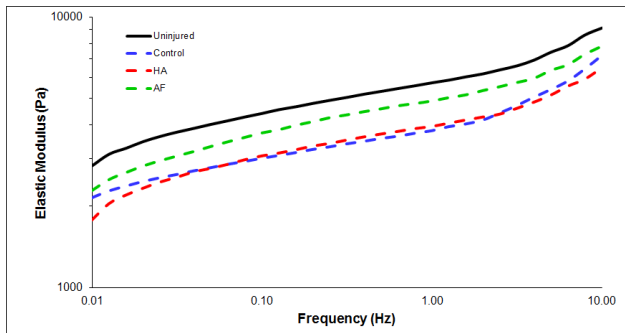
Fig. 1: Rheological analysis showed that animals with biopsy only (control) or biopsy with HA injections had reduced elastic moduli compared with uninjured controls, which also did not appear to differ significantly from each other. Animals receiving the AF injections also showed a reduction in elastic moduli, but to a lesser extent than HA or untreated animals.
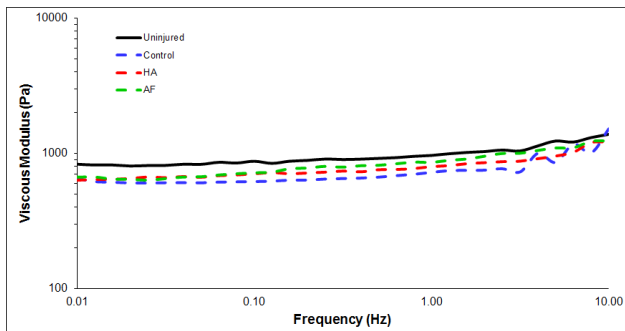
Fig. 2: The viscous moduli for all three experimental groups was well below uninjured values at lower frequencies, with the AF group approaching values of uninjured tissue at higher frequencies.

CONCLUSION

The data show improvement in viscoelastic measurements with AF treatment compared with HA or untreated animals. There was more variability in the AF group than the other groups, with some values comparable to uninjured controls. This suggests AF may be a superior injectable for treatment of vocal fold injury. However, additional studies in this area are necessary and should include a larger cohort as well as adjustment of the dose, concentration, and/or timing of injections.

REFERENCES

[1] Thibeault SL, et al. (2011). In vivo engineering of the vocal fold ECM with injectable HA hydrogels – late effects on tissue repair and biomechanics in a rabbit model. *J Voice*, 25, 249-253

[2] Duflo S, et al. (2006). Effect of a synthetic extracellular matrix on vocal fold lamina propria gene expression in early wound healing. *Tissue Eng*, 12, 3201-3207

[3] Duflo S, et al. (2006). Vocal fold tissue repair in vivo using a synthetic extracellular matrix. *Tissue Eng*, 12, 2171-2180

[4] Hansen JK, et al. (2005). In vivo engineering of the vocal fold extracellular matrix with injectable hyaluronic acid hydrogels: early effects on tissue repair and biomechanics in a rabbit model. *Annals Oto, Rhinol & Laryngol*, 114, 662-670

[5] Rousseau B, et al. (2004). Characterization of chronic vocal fold scarring in a rabbit model. *J Voice*, 18, 116-124

[6] Thibeault SL, et al. (2004). Hyaluronan levels in acute vocal fold scar. *Laryngoscope*, 114, 760-764

[7] Thibeault SL, et al. (2002). Histologic and rheologic characterization of vocal fold scarring. *J Voice*, 16, 96-104

[8] Smith ME, et al. (2016). Wound healing in the airway. Unpublished results.

HUMAN VOCAL-FOLD ARCHITECTURE AND MECHANICAL PROPERTIES: 3D MULTISCALE CHARACTERIZATION AND MODELLING

Thibaud Cochereau[1,2], Alberto Terzolo[1], Lucie Bailly[1], Laurent Orgéas[1], Nathalie Henrich Bernardoni[2], S. Rolland du Roscoat[1], A. McLeer-Florin[3]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, 3SR, Grenoble, F-38000, France
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, F-38000, France
[3]Univ. Grenoble Alpes, CHU Grenoble Alpes, CNRS, Grenoble INP, IAB, Grenoble, F-38000, France

***Keywords*: Vocal folds; Synchrotron X-ray microtomography; Biomechanical tests; Micro-mechanical modelling**

INTRODUCTION

During human phonation, vocal-fold tissues are subjected to complex dynamic loadings involving combined tension, compression and shear solicitations ranging from small to large finite strains. Their complex non-linear and viscoelastic response is closely linked to their 3D multiscale and highly anisotropic fibrous structure. However, the specificities of this structure and how they affect vocal-fold vibro-mechanical properties remain an open question. This is mainly ascribed to the difficulties in properly characterizing the vocal-fold structure [1] and to the lack of available data concerning the mechanics of vocal tissues under various loading conditions. Both limitations restrain the development of suitable 3D mechanical models able to predict, and thus better understand, vocal-fold behavior [1].

Therefore, this study aims at: (i) investigating the 3D multiscale structure of vocal-fold tissues and their mechanical behavior upon finite strains; (ii) proposing a fiber network model based on acquired histological data, able to predict vocal folds non-linear, anisotropic and visco-hyperelastic properties.

METHODS

Human vocal-fold samples

Vocal-fold samples were excised from 15 human larynges dissected within 48 hours *post mortem*. The tissue inner layers, *i.e. lamina propria* (*epithelium* included) and *vocalis* muscle, were subsequently separated on 5 samples.

Vocal-fold structure 2D and 3D imaging

A first set of samples was imaged in 2D using standard histological procedures (Reticulin, Hematoxylin-Eosin-Saffron HES, Masson Trichrome and Elastic stains). A second set, was imaged in 3D using synchrotron X-Ray microtomography with phase retrieval and mean/high spatial resolution (voxel size = 13^3 µm^3 – 0.65^3 µm^3, ID19 beamline of European Synchrotron Radiation Facility, Grenoble, France). The 3D images were then processed to assess the initial orientation distribution, the equivalent diameter D_0 and the waviness ξ_0 of collagen and muscle fibers, respectively in *lamina propria* and *vocalis* layers [2].

Mechanical characterization

Cyclic tension, compression and shear tests at finite strains were conducted at room temperature (25°C) on the third set of samples by means of a dedicated micropress. The device was designed in order to confine the samples in saturated humidity conditions (99% RH) so as to prevent them from tissue drying.

Micromechanical model

A simple micromechanical model was implemented to mimic the 3D structure as well as the mechanical behavior of *lamina propria* and *vocalis* layers. Briefly, these soft tissues are considered as incompressible fiber-reinforced composites. An idealized representative elementary volume (REV) of their fibrous microstructure, similar to the one of the eight chains model developed for rubbers, is shown in the inset of Fig.1a) [3]. It is composed of four wavy (collagen or muscle) fibers, having a volume fraction ϕ and an initial second order fiber orientation tensor $\mathbf{A_0}$. Fibers are thought as wavy fibrils bundles of diameter D_0. The mechanical behavior of folded fibrils is visco-hyperelastic. Knowing the Young's modulus of fibrils E, their initial waviness ξ_0 and diameter d_0, the model accounts for (a) their unfolding (hyperelastic contribution based on the beam model developed in [4]) as well as (b) their viscoelastic interactions with other fibrils and with the surrounding matrix (Maxwell contribution with modulus E_1 and viscosity η_1). The matrix not directly enclosing the fibers is assumed to behave as a Neo-Hookean material with shear modulus μ. Within this framework, the macroscale Cauchy stress tensor $\boldsymbol{\sigma}$ of the tissues is expressed as:

$$\boldsymbol{\sigma} = -p\boldsymbol{\delta} + \boldsymbol{\sigma}_m + \boldsymbol{\sigma}_f, \qquad (1)$$

where p is the incompressibility pressure, $\boldsymbol{\delta}$ the identity tensor, $\boldsymbol{\sigma}_m$ and $\boldsymbol{\sigma}_f$ the stress contributions of the matrix and the fiber network respectively. A rate formulation of $\boldsymbol{\sigma}$ is then used to compute $\boldsymbol{\sigma}$ during the tissue deformation, where, in particular, the local stress rate and stress related to each fiber are estimated in the convected fiber reference frame.

RESULTS AND DISCUSSION

• The 2D histological micrographs shown in Figs.1a) and 2a) convey the typical architectures of collagen and muscle fibers in coronal views within the *lamina propria* and *vocalis* layers, respectively. Collagen fibers exhibit an initial equivalent diameter $D_0 \approx 5$ µm, a pronounced initial waviness $\xi_0 \approx 0.9$ and are initially highly oriented along vocal folds longitudinal direction (*i.e.* antero-posterior direction) $\mathbf{e}_1$ with transverse isotropy:

$$\mathbf{A_0} \approx 0.8\, \mathbf{e}_1 \otimes \mathbf{e}_1 + 0.1\, (\mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3)$$

Vocalis muscular fibers have a larger diameter $D_0 \approx 30$ µm, a limited waviness $\xi_0 \approx 0.96$ and a preferential orientation along $\mathbf{e}_1$ characterized by:

$$\mathbf{A_0} \approx 0.7\, \mathbf{e}_1 \otimes \mathbf{e}_1 + 0.15\, (\mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3)$$
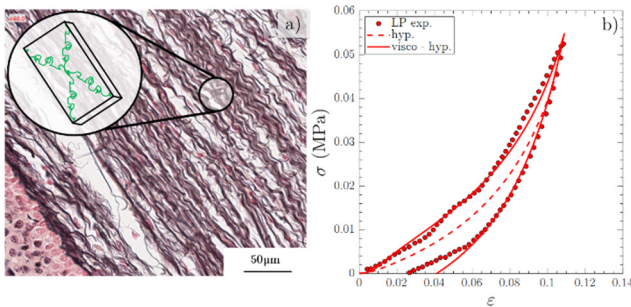
Fig.1: *Lamina propria* **layer: a) typical 2D histological micrograph (Reticulin stain); b) tensile behavior: experimental data versus model predictions.**

• Typical tensile behaviors of *lamina propria* and *vocalis* layers along $\mathbf{e}_1$ are respectively illustrated in Figs.1b) and 2b) (tensile Cauchy stress *vs.* Hencky strain). In both cases, stress-strain curves exhibit non-linear responses with a J-shape strain-hardening (much more pronounced for *lamina propria*), mainly attributed by the unfolding (at least for *lamina propria*) and reorientation of fibers along the tensile direction, stress hysteresis upon unloading and non-negligible residual strain after the unloading, both features related to the viscoelasticity of the considered tissues. It is worth noting that stress levels in *lamina propria* are one order of magnitude higher than those recorded for the *vocalis* layer.

• The predictions of the micromechanical model in tension are also reported in Figs.1b) and 2b), showing how the model provides accurate fits of experimental data for both layers, capturing most of the aforementioned trends. In particular, the role of viscoelastic effects is clearly highlighted (see the model predictions with or without viscoelastic contribution). To obtain these curves, the structural parameters $\mathbf{A_0}, D_0, \xi_0$ estimated from 3D images were used as inputs. Other structural and micromechanical parameters needed for the model were extracted from the literature [5]. For the *lamina propria*, $\phi = 0.45$, $d_0 = 100$ nm and $E = 1000$ kPa were adopted. For the *vocalis*, as available data are scarcer, only $d_0 = 1.5$ µm was used

while ϕ and E were arbitrarily fixed to 0.7 and 75 kPa. In both cases, the shear modulus of the matrix was set to 60 kPa. The viscoelastic parameters E_1 and η_1 were the only unknown parameters of the micromechanical model: they were adjusted in order to fit the experimental stress-strain curves, being equal to 950 kPa and 8 MPa.s, and to 275 kPa and 240 kPa.s for the *lamina propria* and *vocalis* layers respectively.
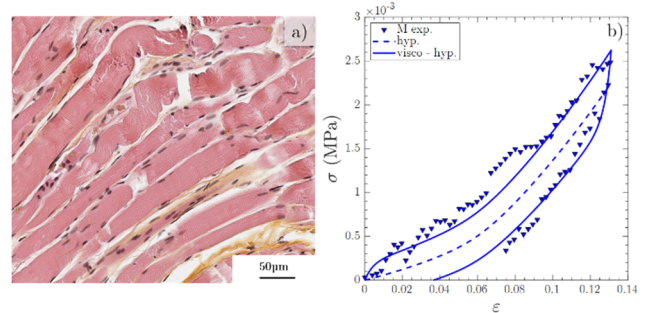
Fig.2: *Vocalis* **layer: (a) typical 2D histological micrograph (HES stain); (b) tensile behavior: experimental data versus model predictions.**

CONCLUSION

Based on 2D histological data and 3D images using synchrotron X-ray microtomography, this study provides a multiscale and quantitative database on the vocal-fold structure at the scale of its muscular and collagen fibers. The role of sublayers' microstructure on their macroscale behavior is assessed through the theoretical predictions of a micromechanical model, successfully compared with experimental mechanical data for tension.

ACKNOWLEDGMENTS

REFERENCES

[1] Miri AK et al. Microstructural characterization of vocal folds toward a strain-energy model of collagen remodeling. *Acta Biomater,* 2013.

[2] Bailly et al., 3D multiscale imaging of human vocal folds using synchrotron X-ray microtomography in phase retrieval mode. *Scientific Reports,* 2018, submitted.

[3] Arruda et al. A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. *J.Mech.Phys.Solids,* Vol.41. No. 2, pp.389-412, 1993.

[4] Bailly et al. Towards a biomimetism of abdominal healthy and aneurismal arterial tissues. *J.Mech.Behavior.Biomed.Mater*, 10:151-165, 2012.

[5] Asgari et al. In vitro fibrillogenesis of tropocollagen type III in collagen type I affects its relative fibrillary topology and mechanics. *Scientific Reports,* 2017.

PHYSIOLOGIC DIFFERENCES AND SUBGROUPINGS FOR NORMAL VOCAL OUTPUT: A DYNAMIC SYSTEMS PERSPECTIVE

Daniel J. Croake[1], Richard D. Andreatta[1], & Joseph C. Stemple[1]

Dept. of Rehabilitation Sciences/Div. of Communication Sciences & Disorders, University of Kentucky, Lexington, KY, USA

***Keywords*: Aerodymanics; Acoustics, Respiratory Kinematics, Dynamic Systems Theory**

INTRODUCTION

Healthy "normal" voice production is thought to arise from the coordinated activity of three component subsystems: *respiration, phonation,* and *resonance* [1-3]. It has been shown that the voice subsystem triad does not function in a unidirectional and linear manner, but rather in a highly interactive fashion in which the subsystems are interdependent, indicating that vocal output is a complex and nonlinear behavioral process resulting from the cooperative interplay of the vocalization subsystems[4, 5].

Normal vocal output is not likely solely dependent on the normality of each individual subsystem measure, but rather the interactive and collective convergence of many potential variables. As such, what constitutes physiologically "normal" vocal output, remains relatively undefined. Because of the interactive complexity among the vocalization subsystems during phonation, a singular characteristic balance, or equilibrium of the subsystem triad is unlikely. A more likely scenario is that normal voice exists on a continuum where multiple collective states of the three subsystems may all result in normal vocal output.

To address this question, the purpose of this study was to (1) characterize and quantify the differences in subsystem patterning in individuals with normal voices via near-simultaneous measurements representative of the vocalization subsystems, and (2) to determine if subgroups could be discerned which further delineate physiologic differences occurring during normal voice production. Lastly, based on these findings, a revised model of phonation is proposed that may provide a more complete picture of the dynamic and interactive interplay of variables within and among the vocalization subsystems.

METHODS

Participants

Twenty-nine healthy young adults with normal voices, 8 men and 21 women, participated. Ages ranged from 20 – 45 years [mean age 26 years (SD = 6.4 years)]. Inclusion criteria were: ages 18 – 50 years; English as native language, perceptually normal speech, language, and voice; and non-smoking for at least the past 5 years. *Exclusion criteria* were: current or prior vocal pathology; stroke or other neurological dysfunction; neck or spinal injury; recent respiratory infections, asthma, or paradoxical vocal fold movement; and formal speaking or singing training.

Instrumentation and Measurement

Vocal fold imaging. Participants received a videostroboscopic exam [Kay PENTAX Rhino-Laryngeal Stroboscope – Model RLS 9100B (Montvale NJ, USA)] coupled to a 70-degree rigid scope (Kay PENTAX Model, SN 1541) to ensure normal vocal function and absence of vocal fold pathology prior to participation.

Acoustic Measurements. Vocal acoustic signals were transduced via omnidirectional condenser microphone [Shure SM-48 (Niles IL, USA)] with a linear frequency response of 55-14kHz. Mouth-to-microphone distance was 3 inches (7.62 cm) at an angle of 45 degrees. The voice signal was digitized at 44.1 kHz and analyzed via spectrography to determine characteristics of the first five vocal formants. Data were analyzed with the PRAAT software package [PRAAT 5381 (Amsterdam, Netherlands)].

Aerodynamic Measurements. Real-time intraoral air pressure and oral airflow data were collected using the Kay PENTAX Phonatory Aerodynamic System (PAS, Model 6000).

Respiratory Measures. Inductance plethysmography was used to measure respiratory kinematic data using the Inductotrace system [Ambulatory Monitoring (Ardsly, NY, USA)] coupled to a Power Lab 16-channel A/D data acquisition system [AD Instruments 16/30 (Colorado Springs, CO, USA)] with data tracings displayed via Lab Chart software ver. 7.2.5 (AD Instruments). Data were stored digitally and analyzed in MATLAB ver.7.4 (Mathworks, Inc.).

Tasks and Conditions. Participants produced ten repetitions of the nonsense sentence, /Buy Pa or Pa a Pa Pa/ [6] at comfortable pitch and loudness, once per breath group. A rate of approximately 1 – 1.5 syllables per second was used. Only the accented (underlined) syllables for each of the ten breath groups were used for analysis, omitting the first and final syllables.

Analysis

Data were analyzed using *JMP* (Cary, NC, USA) by SAS ver.11. Descriptive statistics (means, standard deviations, and ranges) were calculated for all participants. All variables were found to have normal distributions. Hierarchical cluster analysis using the Ward method was

used to determine groups with similar characteristics. Linear Discriminant Analysis (LDA) was then used to determine how well subjects fit into a particular cluster. ANOVA was performed to determine statistical differences of variables between clusters. Post hoc significance for between cluster variables was determined by Tukey's Honestly Significant difference testing (alpha = 0.05). Pearson Product-Moment correlations were performed for all combinations variables in each cluster to determine which variables highlighted interesting features of cluster membership.

RESULTS

Cluster analysis revealed three clusters. LDA determined that all subjects were classified into each of the three clusters without misclassifications. The probability range that participants were properly classified into the established cluster was high ranging from 0.92 – 1.00.
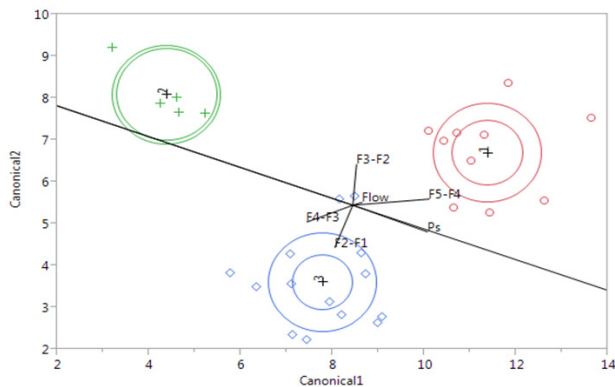
Fig. 1: canonical plot with clusters represented by color. Points on the plot are multivariate means in the two dimensions that best separated the clusters. Circles are 95% confidence intervals for the multivariate means.

ANOVA revealed that seven of the nine variables were found to be significantly different across at least one cluster, indicating that even within the confines of 'normal" voice production, (statistically) significant differences were present among individuals.

DISCUSSION

1. Three distinct clusters were able to be delineated from a larger pool of subjects with normal voices indicating that there are different physiologic mechanisms which result in normal voice production. Seven of nine measured variables were found to be significantly different across at least one cluster.

2. Significant correlations were observed within each cluster that revealed important interactive relationships within subsystems and indicate not only that individuals are different, but how they differ from one another.

3. Subglottic pressure was a variable common to all clusters, potentially indicating a key role in the ordered

emergence of voicing as a potential control parameter regardless of the subgroupings of individuals.

CONCLUSION

Normal vocal output does not appear to be produced in the same manner among all individuals; however it does not appear to be completely idiosyncratic either. Three distinct and preferred performance patterns were observed which revealed differences in the interactive nature of the subsystem triad. In the context of dynamic systems, subgroups may represent preferred functional states which are stable performance patterns (attractors) in certain individuals. Understanding the nature of these patterns will help to develop more individualized and focused treatments for patients with voice disorders.
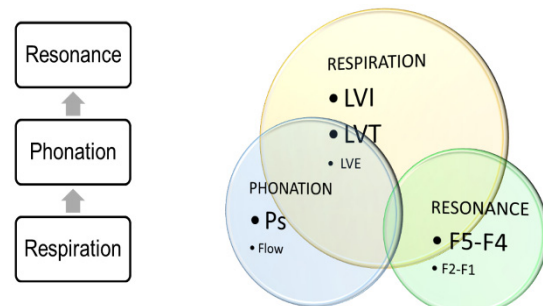
Fig. 2: Traditional subsystem model (left) and dynamic model (right). The dynamic model is more flexible to account for the individual differences that exert a performance bias among the subsystem triad.

REFERENCES

1. Isshiki, N., *Regulatory mechanism of voice intensity variation.* Journal of Speech and Hearing Research, 1964. **7**: p. 17-29.

2. Kent, R.D., *The speech sciences.* 1997, San Diego, CA: Singular.

3. Stemple, J., L. Glaze, and B. Klaben, *Clinical voice pathology: Theory and management.* 3rd ed. 2000, San Diego: Singualr Publishing group.

4. Titze, I.R., *Nonlinear source-filter coupling in phonation: theory.* J Acoust Soc Am, 2008. **123**(5): p. 2733-49.

5. Huber, J.E. and E.T. Stathopoulos, *Respiratory and laryngeal responses to an oral air pressure bleed during speech.* J Speech Lang Hear Res, 2003. **46**(5): p. 1207-20.

6. Stathopoulos, E.T., et al., *Increased vocal intensity due to the Lombard effect in speakers with Parkinson's disease: simultaneous laryngeal and respiratory strategies.* J Commun Disord, 2014. **48**: p. 1-17.

EMERGENT INTERACTIONS AMONG THE VOCALIZATION SUBSYSTEM TRIAD: A DYNAMIC SYSTEMS APPROACH

Daniel J. Croake[1], Richard D. Andreatta[1], & Joseph C. Stemple[1]

Dept. of Rehabilitation Sciences/Div. of Communication Sciences & Disorders, University of Kentucky, Lexington, KY, USA

***Keywords*: Aerodynamics; Acoustics, Respiratory Kinematics, Dynamic Systems Theory**

INTRODUCTION

The idea that there are individual differences in voice production is not new[1]; however it would be clinically important to better understand the features of these differences in order to provide more personalized treatments for voice disorders. Voice production is thought to emerge from a relative balance of variables within and among the vocalization subsystems of respiration, phonation, and resonance; however only a handful studies specifically report on the interactive relationships among the three subsystems. Because voice production is known to be a nonlinear process[2], we adopted Dynamic Systems Theory (DST) as an experimental framework from which to study how individuals cope with a perturbation to the larynx in the form of a temporarily induced unilateral vocal fold paralysis (UVFP).

Briefly, a key concept of DST is that the emerging order of component elements forming a functional system tend to collect themselves into loosely stable patterns and interactive states known as *attractors*[3]. Attractors are a theoretical construct that characterize the preferred performance states of the system's elements including its stability to disruption and range of acceptable performance. If an attractor is sufficiently disturbed, or if the range of acceptable behavioral performance is exceeded, the system seeks out a new stable performance state in which it can function efficiently. In voice production, such a scenario suggests that although a theoretically infinite possibility of performance states exist (what is referred to in DST as the "state space"), only a few collective states will be "preferred," signifying that there are likely only a relative few observable performance patterns that characterize what we recognize as normal (or disordered) vocal output.

DST differs critically from reductionist approaches in its theoretical recognition that the performance characteristics of a complex system cannot be entirely explained by the action of only one or few of the complex system's component parts. Rather, it is necessary to understand how these separate elements function interactively within a given task and environmental context [4]. A key means of investigating the nature of a complex system is therefore accomplished by perturbing the system and observing how the system responds, compensates, and restores itself back to its original performance condition.

The aim of this study was to quantify the interactions of the three vocalization subsystems before, during, and after a perturbation to the larynx (temporarily lidocaine induced UVFP) in 10 vocally healthy participants. Using DST as a guide, we hypothesized that data groupings would emerge revealing context dependent patterns in the relationships of variables representing the three vocalization subsystems. We also hypothesized that group data would mask important individual variability important to understanding the relationships among the vocalization subsystems.

METHODS

Respiratory kinematic, aerodynamic, and acoustic formant measures were obtained from 10 healthy participants (8F, 2M) with normal voices before, during, and after a temporarily lidocaine induced UVFP. Inclusion criteria were: ages 18 – 50 years; English as native language, perceptually normal speech, language, and voice; and non-smoking for at least the past 5 years. *Exclusion criteria* were: current or prior vocal pathology; stroke or other neurological dysfunction; neck or spinal injury; recent respiratory infections, asthma, or paradoxical vocal fold movement; formal speaking or singing training; and allergies to Lidocaine, Epinephrine, or its derivatives.

Instrumentation and Measurement

Vocal fold imaging. Participants received a videostroboscopic exam [Kay PENTAX Rhino-Laryngeal Stroboscope – Model RLS 9100B (Montvale NJ, USA)] coupled to a 70-degree rigid scope (Kay PENTAX Model, SN 1541) prior to measurements in all three conditions to document vocal fold mobility.

Acoustic Measurements. Acoustic signals were transduced via omnidirectional condenser microphone [Shure SM-48 (Niles IL, USA)] with a linear frequency response of 55-14kHz. Mouth-to-microphone distance was 3 inches (7.62 cm) at an angle of 45 degrees. The voice signal was digitized at 44.1 kHz and analyzed via spectrography to determine characteristics of the first five vocal formants. Data were analyzed with the PRAAT software package [PRAAT 5381 (Amsterdam, Netherlands)]

Aerodynamic Measurements. Real-time intraoral air pressure and oral airflow data were collected using the Kay PENTAX Phonatory Aerodynamic System (PAS, Model 6000).

Respiratory Measures. Inductance plethysmography was used to measure respiratory kinematics using the Inductotrace system [Ambulatory Monitoring (Ardsly, NY, USA)] coupled to a Power Lab 16-channel A/D data acquisition system [AD Instruments 16/30 (Colorado Springs, CO, USA)] with data tracings displayed via Lab Chart software ver. 7.2.5 (AD Instruments). Data were stored digitally and analyzed in MATLAB ver.7.4 (Mathworks, Inc.).

Tasks and Conditions. Participants produced ten repetitions of the nonsense sentence, /Buy Pa or Pa a Pa Pa/ [5] at comfortable pitch and loudness, once per breath group. A rate of approximately 1 – 1.5 syllables per second was used. Only the accented (underlined) syllables for each of the ten breath groups were used for analysis, omitting the first and final syllables. Tasks were completed before, during, and after recovery from the UVFP.

Analysis

Data were analyzed using *JMP* (Cary, NC, USA) by SAS ver.11. Descriptive statistics (means and standard deviations) were calculated for all participants. All variables were found to be normally distributed. Group differences across task conditions were assessed with repeated-measures ANOVA. Tukey's Honestly Significant Difference (HSD) testing was applied to determine significant differences for variables across conditions (alpha = 0.05). Correlations were calculated to determine relationships among the formant spacings (alpha = 0.05). All dependent variables were standardized with z-scores so that all variables could be compared on the same scale.

RESULTS & DISCUSSION

1. Resonance changes during the paralytic condition were characterized by a narrowing of the spacing of the upper formants and a widening of the spacing of the lower formants. This is important as format spacing relates to a physiologic change in pharyngeal shape and suggests epilaryngeal narrowing. Physiologically, some degree of supraglottic constriction in the epilarynx area was noted for all subjects in the DUR condition on visualization.

2. Despite similar positions of the paralytic vocal fold (paramedian in all cases) very different alterations in airflow rate resulted presumably from the differing compensatory strategies used across subjects.

3. Subglottic pressure was generally maintained across all conditions indicating that a primary compensatory strategy in response to glottic incompetence may be maintenance of subglottic pressure from the resulting increase in airflow rates.

4. Respiratory strategy varied greatly by subject. A more varied and active respiratory response was generally observed with greater increases in airflow rate. All but one participant had lower lung volume initiations

indicating that taking advantage of passive recoil properties of the respiratory system was not a preferred response during the laryngeal perturbation.

5. Ribcage and abdominal proportional contributions varied markedly in response to increased airflow rate becoming more variable with greater airflow change.

6. Not all subjects returned to their baseline subsystem vocalization strategy in the recovery phase. This may be due to the fact that measures were taken only a few minutes after confirming returned vocal fold mobility. Another possibility is that the paralytic state may function as a stable attractor pattern of its own that requires greater time to dissipate.

CONCLUSION

Participants appeared to compensate to the laryngeal perturbation in different ways, but could be grossly categorized into smaller subgroupings with similar characteristics. Group data alone masked important individual variability necessary to understand the unique relationships among the three subsystems. Multi-level analysis permitted a richer understanding of the individual differences in phonatory regulation and permitted sub-group analysis. DST may be a useful heuristic to model the interactive relationships among vocalization subsystems; however future data with larger subject pools will be necessary in order to draw firm conclusions.

REFERENCES

1. Isshiki, N., *Regulatory mechanism of voice intensity variation.* Journal of Speech and Hearing Research, 1964. **7**: p. 17-29.

2. Titze, I.R., *Nonlinear source-filter coupling in phonation: theory.* J Acoust Soc Am, 2008. **123**(5): p. 2733-49.

3. Prigogine, I. and I. Stengers, *Order out of chaos: man's new dialogue with nature.* 1984, New York, NY: Bantam Books.

4. Thelen, E. and B.D. Ulrich, *Hidden skills: a dynamic systems analysis of treadmill stepping during the first year.* Monogr Soc Res Child Dev, 1991. **56**(1): p. 1-98; discussion 99-104.

5. Stathopoulos, E.T., et al., *Increased vocal intensity due to the Lombard effect in speakers with Parkinson's disease: simultaneous laryngeal and respiratory strategies.* J Commun Disord, 2014. **48**: p. 1-17.

EXACT ANALYTICAL MODEL BASED ON THE THEORY OF CONFINED FLOWS FOR THE ASSESSMENT OF THE RELATIONSHIP BETWEEN TUBE PARAMETERS AND TOTAL FLOW RESISTANCE IN SEMI-OCCLUDED VOCAL TRACT EXERCISES

Andrey Ricardo da Silva[1], Ana Carolina Ghirardi[2], Matheus R. Reizer[1], Stephan Paul[1]

[1] Department of Mechanical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil
[2] Speech-Language Pathology and Audiology Department, Federal University of Santa Catarina, Florianópolis, SC, Brazil

Keywords: Semi-occluded vocal tract exercises; Resonance tube phonation; Back pressure; Flow Resistance

INTRODUCTION

Semi-occluded vocal tract exercises (SOVTEs) are widely used in clinical settings for improving source-filter interaction and impedance adjustments of the vocal tract with excellent results. These exercises increase oral cavity pressure (back pressure), which act to augment vocal fold adduction. Furthermore, the decrease in transglottal pressure pushes the vocal processes slightly apart. Therefore, these techniques provide patients with important sensory feedback, promoting greater vocal economy and reducing the amplitude of vocal fold vibration [1-3].

Among the several different SOVTEs, techniques involving phonation into straws or tubes of different materials, lengths, and diameters are highly beneficial in the prevention and treatment of voice disorders. Short plastic stirring straws or longer drinking straws may be used with their free end in air or immersed in water for increased resistance [4,5].

Variations in tube wall roughness, length, and inner diameter, as well as different free-end conditions during exercise execution (tube with its free end in air or immersed in water) will drastically modify the tube's flow resistance, the total pressure drop (back pressure) and, ultimately, the mechanisms associated with the effects on the voice, as observed in patients [4, 6,7].

Although clinical studies have been successful at addressing the effectiveness of SOVTEs on improving voice parameters, they lack substantial information on how these improvements are achieved from a physical point of view. Furthermore, studies that address the influence of geometric properties of the tube on flow resistance are still sparse in the literature.

The purpose of this study is to demonstrate an exact analytical model, based on the theory of confined flows, for the assessment of the relationship between length, inner diameter and flow resistance in tubes used for SOVTEs.

METHODS

The underpinning reference in this study is the Darcy-Weisbach equation for confined flows. As the back pressure generated by SOVTEs is the sum of the various pressure drop mechanisms along the tube, equations representing the different pressure drop mechanisms were combined and analyzed, illustrating the influence of the tubes geometric parameters on the back pressure and flow resistance. Based on these equations, an isometric resistance chart was plotted, showing the relationship between length and inner diameter of any given tube and flow resistance.

RESULTS AND DISCUSSION

The analysis and combination of the equations representing the pressure drop mechanisms involved in the back pressure provided by SOVTEs, i.e. viscous loss, kinetic energy loss, and loss due to the water column yielded Equation (1), where k is the sum of the loss coefficients at the two terminations of the tube.

$$p_{back} = U^2 \frac{8\rho}{\pi^2}\left[f_{DW}\frac{L}{D^5} + \frac{k}{D^4} \right] + \rho_w g h. \tag{1}$$

Equation (2) is derived from the expression above, and illustrates flow resistance as the ratio between the pressure drop Δp_T (back pressure) and volume flow U.

$$R = U \frac{8\rho}{\pi^2}\left[f_{DW}\frac{L}{D^5} + \frac{k}{D^4} \right] + \frac{\rho_w g h}{U} \tag{2}$$

Overall, the results obtained from the model agreed well with both experimental and theoretical results provided by the literature, considering different geometries and free-end conditions (tubes with the free end in air or immersed in water).

Particularly for tubes with their free end in air, the results agree with those reported by Smith and Titze [8] who tested tubes of different lengths and diameters, especially for tubes that are longer than 6cm, with diameters greater than 4.1mm. As the tube becomes shorter and thinner, the back pressure (p_{back}) becomes more

pronounced. In these cases, Equation (1) slightly underestimated the p_{back}, with a maximum error of $\approx 5\%$, which would be acceptable for practical purposes when using short and thin tubes.

For tubes with one end immersed in water, the agreement of the results provided by Equation (1) with previous experimental reports in literature depended on the height of the water column. Results for higher water depths (h>4cm) agree better with the findings of Andrade et al. [9] whereas shorter immersions agreed well with the results reported by Smith and Titze [8].

Based on Equation (2), an isometric resistance chart was plotted (Figure 1), so that clinicians may find an equivalent replacement for a tube of known length (L) and diameter (D), providing the same flow resistance in SOVTEs.
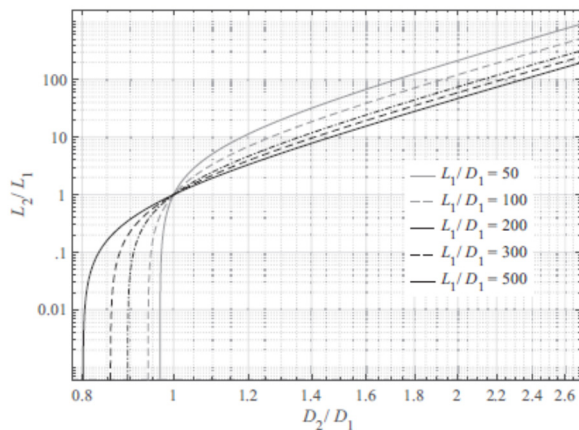
Fig. 1: Isometric resistance chart. L_1, D_1 are the length (mm) and diameter (mm) of the reference, respectively. Likewise, L_2, D_2 are the length (mm) and diameter (mm) of the equivalent tube, respectively.

For instance, assuming a reference tube to have L1= 250 mm and D1= 5 mm, one wants to find the necessary length L2 of an equivalent tube of diameter D2=8 mm as to provide the same flow resistance given by the reference tube. In this case, the ratio between equivalent and reference diameters would be D2/D1= 1.6 and the ratio between length L1 and diameter D1 of the reference tube would be 50. For those values, the chart would give the length ratio L2/L1=70, which implies that the length of the equivalent tube L2 would have to be 250mm·70=17.5m in order to produce the same flow resistance provided by the reference tube.

Thus, even very small variations on the tube diameter require huge variations on the tube's length to maintain flow resistance.

CONCLUSION

Flow resistance is significantly more sensitive to variations in the inner diameter than to the tube's length. Furthermore, viscous losses are much more significant than kinetic energy losses for tubes whose length is at least one order of magnitude greater than the inner diameter. For short tubes (L/D<10), flow resistance will be governed mainly by the ratio between the tube's inner diameter and the effective diameter of the patient's mouth. Finally, the study proposes an isometric resistance chart that may be used to indicate the dimensions of an equivalent tube, which would provide the same flow resistance produced by a reference tube, which may be useful for selecting appropriate tubes for use in SOVTEs in clinical settings.

REFERENCES

[1] Titze I. Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings, *J Speech Lang Hear Res*. 2006;49:448–459.

[2] Guzman M, Laukkanen A, Krupa P, et al. Vocal tract and glottal function during and after vocal exercises with resonance tube and straw. *J Voice*.2013;27:523–534.

[3] Berry D, Verdolini K, Montequin D, et al. A quantitative output-cost ratio in voice production. *J Speech Lang Hearing Res*. 2001;44:29–37.

[4] Wistbacka G, Andrade P, Simberg S, et al. Resonance tube phonation in water—the effect of tube diameter and water depth on back pressure and bubble characteristics at different airflows. *J Voice*. 2018;32:126.e11–126.e22.

[5] Tyrmi J, Radolf J, Horacek J, et al. Resonance tube or lax vox? *J Voice*. 2017;31:430–437. in press.

[6] Mills R, Rivedal S, DeMorett C, et al. Effects of straw phonation through tubes of varied lengths on sustained vowels in normal-voiced participants. *J Voice*. 2017;in press. doi: 10.1016/j.jvoice.2017.05.015

[7] Simberg S, Laine A. The resonance tube method in voice therapy: description and practical implementations. *Logoped Phoniatr Vocol*. 2007;32:165–170.

[8] Smith S, Titze I. Characterization of flow-resistant tubes used for semioccluded vocal tract training and therapy. *J Voice*. 2017;31:113-121.

[9] Andrade PA, Wistbacka G, Larsson H, et al. The flow and pressure relationships in different tubes commonly used for semi-occluded vocal tract exercises. *J Voice*. 2016;30:36–41.

THE INFLUENCE OF HIGH-SPEED VIDEOENDOSCOPY DATA QUALITY ON REDUCED-ORDER MODEL PARAMETERS ESTIMATED USING BAYESIAN INFERENCE

Jonathan J. Deng, Paul J. Hadwin, Sean D. Peterson[*]

Department of Mechanical and Mechatronics Engineering, University of Waterloo, Waterloo, Ontario, Canada
*Corresponding author: peterson@mme.uwaterloo.ca

Keywords: Bayesian Inference; Inverse Analysis; High-Speed Videoendoscopy; Reduced-Order Modelling

INTRODUCTION

Models of phonation present a potentially powerful tool to aid clinicians in the diagnosis of voice disorders, as they allow for estimating factors that are difficult to measure in the clinic [1]. Due to the difficulty of establishing patient specific model parameters however, models are often restricted to population averaged predictions. Inverse analysis techniques have recently been used to overcome this by combining the clinical and modelling approaches, inferring patient specific model parameters from direct clinical measurements [1, 2, 3].

Recently, high-speed videoendoscopy (HSV) of the vocal folds (VFs) has been shown to be a data-rich measurement for parameter inference [2, 3]. However, these previous works, for simplicity, either ignored or employed a simple estimation of uncertainty inherent in HSV. In particular, HSV is sensitive to many variables, such as the camera viewing angle, image resolution, frame rate, and lighting conditions. In this work, we aim to quantify the effects that viewing angle, image resolution, and frame rate have on estimates of stationary parameters computed from HSV using Bayesian inference [1].

METHODS

Instrumentation and Measurement

A simulated HSV recording of driven synthetic VFs was used to estimate subglottal pressure, P_{sub}, and cricothyroid muscle activation, a_{ct}, of a reduced order model of the VFs [4]. HSV was simulated by recording the motion of two dimensional symmetric synthetic rigid VFs, shaped according to the M5 geometry of Scherer *et al.* [5], using a Nikon D3200 camera. The motion of the synthetic VFs was controlled using two stepper motors for each VF, one for rotation and one for lateral translation. The position of each VF was determined by a three mass body cover model (BCM) parameterized using muscle activation rules presented by Titze and Story, and driven by a one-dimensional Bernoulli flow [4].

In order to simulate HSV, the VFs were scaled up by a factor of 7.5 and the motion was slowed down by a factor of 2800, resulting in an effective frame rate and resolution of 84000 FPS and 750 px mm^{-1}, respectively. The effects of frame rate and resolution were then studied by downsampling the captured videos. Viewing angle was examined by physically rotating the camera to provide an off-axis view of the driven VF models. The experimental parameters investigated are given in Table 1.

Table 1: HSV imaging variables tested.

Variable	Values
Angle of view [°]	0.0, 2.5, 5.0, 7.5, 10.0
Frame rate [FPS]	84000, 10500, 5250, 2625
Resolution [px mm^{-1}]	750, 188, 94, 47

Analysis

Videos were captured for four parameter sets ((a_{ct}, P_{sub}) = (0.15, 1800 Pa), (0.15, 2000 Pa), (0.20, 1800 Pa), and (0.20, 2000 Pa)) of the BCM at each of the viewing angles listed in Table 1. Video was recorded for 7.5 minutes, corresponding to approximately 16 complete vibration cycles. A glottal width waveform (GWW) was computed for each video using a zero-crossing-based Laplacian edge detector.

Estimates of P_{sub} and a_{ct} were computed using Bayesian inference. Bayesian inference centers on Bayes' formula

$$\pi_{post}(\mathbf{x}|\mathbf{y}) = \frac{\pi_{like}(\mathbf{y}|\mathbf{x})\pi_{pri}(\mathbf{x})}{\pi_{ev}(\mathbf{y})}, \quad (1)$$

where $\mathbf{x} = [a_{ct}, P_{sub}]^{\mathrm{T}}$, is a vector of patient specific model parameters to be inferred, $\mathbf{y}$ is the GWW, and π represents a probability density [6]. The term $\pi_{post}(\mathbf{x}|\mathbf{y})$ is the posterior probability density and represents all known statistical information about the parameters, x, based on the observations, $\mathbf{y}$, and any *a priori* knowledge of the parameters.

The likelihood relates the residual between the model and the data, and describes the chance of observing the data, $\mathbf{y}$, given the system configuration defined by $\mathbf{x}$, in the context of noise and model error. Specifically, when the measurement noise is treated as an unbiased and identically distributed normal random variable, it is given by

$$\pi_{like}(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi\sigma^2)^k}} e^{\frac{-1}{2\sigma^2}\|\mathbf{y}-F(\mathbf{x})\|^2}, \quad (2)$$

where σ is the standard deviation of the measurement noise (set to 10% of the GWW amplitude herein), k is the number of measurements in $\mathbf{y}$, and $F(\mathbf{x})$ is the forward model that computes the GWW using the BCM for given values of a_{ct} and P_{sub} [6]. Estimates of a_{ct} and P_{sub} were computed by maximizing the likelihood, called maximum likelihood estimates (MLE), which correspond to the traditional least squares estimate. Normalizing the likelihood so it integrates to unity however, allows for computation of the variance, which is used as an estimate of the uncertainty in the MLE [6].

RESULTS AND DISCUSSION

The effects of frame rate, resolution, and angle of view on the estimates are examined through their effects on the MLE estimate and relative variance (the ratio of variances in the non-downsampled video and orthogonal angle of view to the modified video). Here only a single parameterization (a_{ct}=0.15, P_{sub}=1800) is shown for brevity, although all other cases present similar trends.

As the frame rate is reduced, a proportional increase in the parameter uncertainty is observed, while the mean estimate remains constant, as seen in Figure 1.
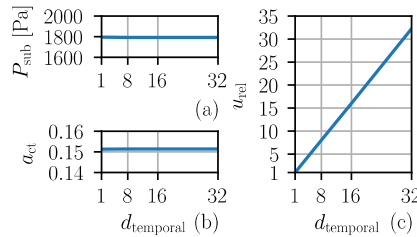
Figure 1: Trends in maximum likelihood (a), (b) and uncertainty (c) with decreasing framerate.

Similar to how repeating measurements of an object decreases uncertainty, reducing the number of frames (or measurements) reduces the certainty on the estimated parameters in this case. In contrast, even the largest downsampling factor (2625 FPS) does not produce a significant bias in the estimate. This is because the frame rate is still large enough to capture the shape of the GWW. Under further downsampling this would not hold, due to aliasing of the GWW.
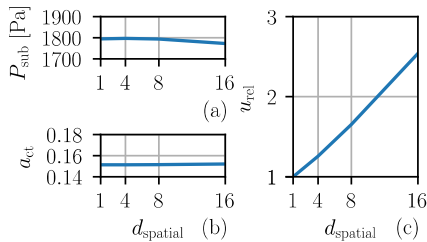
Figure 2: Trends in maximum likelihood (a), (b) and uncertainty (c) with decreasing resolution.

Figure 2 shows the trends as the resolution was downsampled. We see a linear increase in the relative uncertainty, and a slight bias in the estimated parameters. The increase in the relative uncertainty is a result of decreasing precision with decreasing resolution in the edge detection. The parameter σ in the likelihood was scaled according to this increase, leading to the observed increase in relative uncertainty. The bias in the estimated parameters was due to a bias in the edge detection procedure, which underestimated the glottal width at lower resolutions. This resulted in a decrease in the P_{sub} estimate. The exact reason for the bias requires further investigation; other edge detection algorithms may avoid a bias. Compared to an equivalently downsampled frame rate, the downsampled resolution produces smaller increases in uncertainty, which suggests that resolution can be compromised in favor of increased frame rate to reduce uncertainty in estimated parameters from HSV.

As the angle of view is changed, there is a significant bias in the estimated parameters and an increase in the uncertainty, as seen in Figure 3.
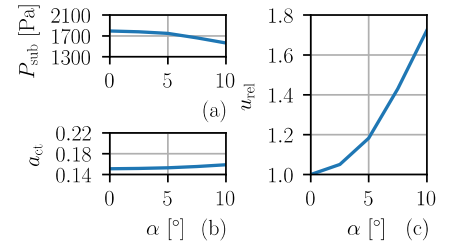
Figure 3: Trends in maximum likelihood (a), (b) and uncertainty (c) with decreasing angle of view.

The bias in the estimated parameters is due to the reduction in the measured glottal width as a result of the parallax effect with off-axis viewing. Since the measurement noise model did not account for this error, the extents of the probability distribution do not take into account the angle of view, and are not captured in the model.

CONCLUSION

The effect of three HSV imaging variables on Bayesian inference were investigated. Frame rate and resolution primarily affected uncertainty of the estimated parameters, with frame rate having a stronger influence than resolution. This suggests that framerate can be increased at the expense of resolution to improve certainty on estimates. In contrast, the angle of view was found to significantly influence the estimated parameter values. In real recording cases, where the angle of view is not easily controlled or determined, it should be modeled as an uncertain parameter.

REFERENCES

[1] Hadwin PJ, Galindo GE, Daun KJ, Zañartu M, Erath BD, Cataldo E, Peterson SD. Non-stationary Bayesian estimation of parameters from a body-cover model of the vocal folds. *J Acoust* 2016; 139(5):2683-2696.

[2] Schwarz R, Döllinger M, Wurzbacher T, Eysholdt U, Lohscheller J. Spatio-temporal quantification of vocal fold vibrations using high-speed videoendoscopy and a biomechanical model. *J Acoust* 2008; 123(5):2717-2732.

[3] Cataldo E, Soize C, Sampaio R. Uncertainty quantification of voice signal production mechanical model and experimental updating. *Mech Syst Signal Process* 2013; 40(2):718-726.

[4] Titze I, Story B. Voice simulation with a body-cover model of the vocal folds. *J Acoust 1995*; 97(2):1249-1260.

[5] Scherer RC, Shinwari D, De Witt KJ, Zhang C, Kucinschi BR, Afjeh AA. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees. *J Acoust* 2001; 109(4):1616-1630.

[6] Kaipio J, Somersalo E. *Statistical and Computational Inverse Problems*. Springer Science and Business Media, 2006.

PRIMO PASSAGGIO: MEASURES ASSOCIATED WITH DIFFERENT INTERPRETATIONS SUNG BY AN ELITE SOPRANO PERFORMER

Raymond Diaz[1], Ronald Scherer[1], Jennifer Rowley

[1] Communication Sciences and Disorders, Bowling Green State University, Bowling Green, Ohio, U.S.A.

***Keywords*: Voice; Voice Science; Primo Passaggio; Emotion**

INTRODUCTION

The primo passaggio is the sequence of notes in a singer's voice that typically differentiates the chest voice from the head voice. Negotiating the primo passaggio is an important goal for classical singers to attain virtuosity. Register transitions can be used to signify a character's musico-dramatic intention. The primary aim of this study was to observe the acoustic and aerodynamic adjustments that an elite soprano performer makes when vocally creating contrasting interpretations of the same lines from the standard operatic canon. This was an exploratory study with a single participant, a soprano who currently performs at the Metropolitan Opera in NYC. Short excerpts from two operas were the focal points of the study. The excerpts were sung to show contrast in intentional registration shifts through the primo passaggio. Analyses were of simultaneously recorded wideband airflow, microphone, and EGG signals. The two performance interpretations for each excerpt varied the perceived amount of chest and head registration according to self-report by the performer. This exploratory study hypothesized that the contrasting interpretations would yield distinctly different objective information (acoustic, aerodynamic, glottographic), and thus be pedagogically important for artistic training.

METHODS

Human Data

The study had a single participant, Jennifer Rowley (she has given permission to use her name). Since the time of the study, Ms. Rowley has sung the operas *Il Trovatore* and *Tosca* (from which the excerpts of the study were selected) at the Metropolitan Opera, Paris National Opera, and the Philadelphia Orchestra, and at other operatic houses.

Ms. Rowley was asked to make contrasting interpretations of "…soghigno di demone!" from Puccini's *Tosca* (sung by the titular character) and of "…che l'anima tutta mi vede!" from Verdi's *Il Trovatore* (sung by Leonora). For each of the excerpts, she was asked to sing an interpretation that was "aside" or "indirect" and to contrast it with a more "aggressive" or "confrontational" interpretation.

Instrumentation and Measurement

The singing was recorded in a sound treated booth (IAC Model 402A) in the Voice Laboratory at BGSU. The participant completed a consent form and questionnaires prior to the recording to detail her recent performance roles, recent medical history, voice *fach*, current teacher, and normal weekly practice time (she was healthy at the time of recording). Simultaneous recordings were made of the microphone signal (Model AKG C 420), at a mouth-to-microphone distance of approximately 7 cm, the signal from an electroglottograph (Kay Elemetrics), and aerodynamic airflows and oral air pressures obtained via a circumferentially ventilated flow mask (Glottal Enterprises, model MSIF-2 S/N). Digitization was performed by an A/D converter (DATAQ Instruments, model DI-2108, with WINDAQ software) outside the sound booth.

Analysis

The recordings were analyzed using Matlab®-based Sigplot custom software, Praat software (version 6.0.26), and Microsoft Excel. Fundamental frequency, intensity, and airflow data were exported to Excel files from Sigplot. The airflow was smoothed (averaged) using a weighted 7-sample approach. The electroglottographic signal was analyzed in Sigplot for the EGGW50 measure, a relative pulse width measure at the 50% height location on the EGG waveform that is first altered to remove any DC shift.
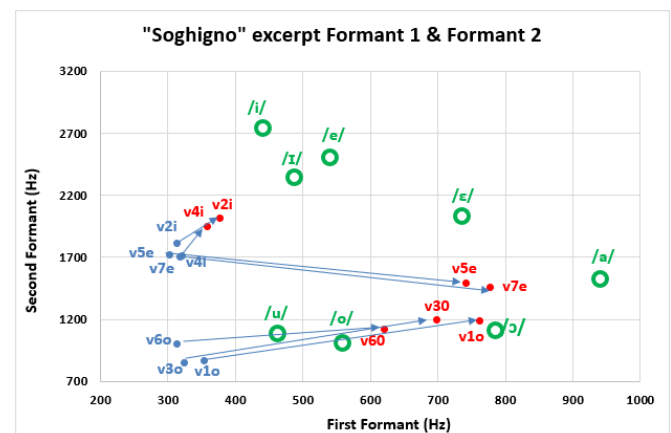
Fig. 1: First and second formant values for the sequential vowels of the two interpretations of the phrase "Soghigno di demone!".

RESULTS

In Figure 1, each point represents the formant values of individual vowels within the phrase "Soghigno di demone!" The blue symbols correspond to the "Indirect Condemnation" interpretation, and the red symbols correspond to the "Confrontational Hatred" vowels. The designations are vowel "v", sequential number of the vowel in the phrase "v1", followed by the vowel itself "v1o". The green circles are the formant values given by Hillenbrand *et al.* (1995) [2] from Table 7-1 in Baken and Orlikoff (2000) [1].

In figure 2, airflow markedly drops in the second half of the word "tutta" in the "provocative" interpretation of the phrase "che l'anima tutta mi vede!" It is in the same moment that the singer perceived a shift through the primo passaggio into a near complete (90%) chest voice. In the study, these data are contrasted by the "aside" interpretation, where the airflow remained constant throughout the same phrase.

The main findings were that:

(1) large formant changes for the intended vowels of the excerpts accompanied the interpretation choices, resulting in large spectral differences for the "same vowels" in the score, where lower first formant values corresponded to the "indirect" interpretations, in contrast to the higher first formant values for the "confrontational" interpretations as shown in Fig. 1;

(2) the more "provocative" interpretations were sung in less time;

(3) vibrato extent and rate were varied selectively for specific vowels, but on average were similar between interpretations;

(4) typically the average airflow was greater in the interpretations that were considered more "indirect" and would be produced "aside" on stage;

(5) the more provocative interpretations had greater airflow spikes (peaks) relative to consonant production, suggesting subglottal pressure control;

(6) the primary example of a registration shift from head to chest register was accompanied by a severe drop in airflow suggesting a significant rise in laryngeal airflow resistance (and glottal adduction); and

(7) a primary indication of intent for one of the interpretations was a large variation in the intensity for emphasis.

DISCUSSION AND CONCLUSIONS

The project observed, recorded, and analyzed the acoustic, aerodynamic, and electroglottographic data gathered in the singing of standard repertoire of Western Opera. It was done in hopes of increasing the corpus of knowledge of both voice science and vocal pedagogy.
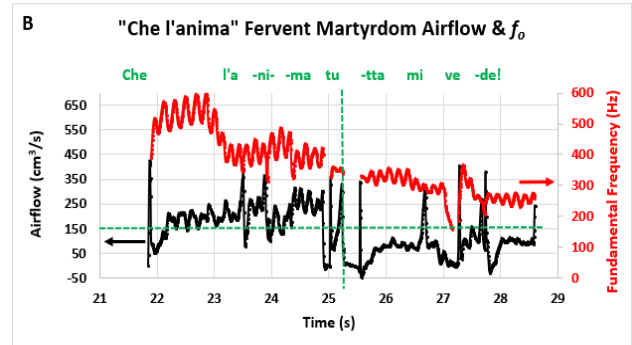
Fig. 2: Simultaneous averaged airflow (black, lower trace) and fundamental frequency (red, upper trace) for "che l'anima tutta mi vede!" excerpt. B: "Fervent Martyrdom" interpretation.

The research suggests that there is a need to study interpretation consequences on the objective measures in singing, the findings giving greater understanding of basic processes of phonation and communication intent, and basic pedagogical information to guide singers to valid interpretation practices.

REFERENCES

[1] Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice.* 2nd ed. San Diego, CA: Singular Publishing Group; 2000.

[2] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97*(5 Pt 1), 3099.

ADDUCTORY VOCAL FOLD KINEMATIC TRAJECTORIES DURING CONVENTIONAL SPEED VS. HIGH-SPEED VIDEOENDOSCOPY

Manuel Diaz-Cadiz[1], Victoria S. McKenna[1], Jennifer Vojtech[1,2], Cara E. Stepp[1,2,3]

[1] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA
[2] Department of Biomedical Engineering, Boston University, Boston, MA, USA
[3] Department of Otolaryngology – Head and Neck Surgery, Boston University, Boston, MA, USA

Keywords: e.g. vocal fold kinematics; high-speed videoendoscopy.

INTRODUCTION

The purpose of this study was to examine vocal fold (VF) adductory angle trajectories using high-speed videoendoscopy (HSV) to determine whether VF adductory behavior is sigmoidal (suggested by previous work using low-speed sampling of 30 fps) and whether sampling rate affects the appropriateness of sigmoidal fits to VF trajectories.

METHODS

Twenty-six participants with healthy voices (10 males; M 20.9 years, SD 2.8 years) underwent flexible HSV at 1000 fps, performing /ifi/ sequences at different speech rates and effort levels. A semi-automatic technique was developed to extract VF angle waveforms during adduction from HSV recordings. VF angles from videos of sufficient quality (1731 cases) were extracted and down-sampled to simulate different lower frame rate conditions (30, 60, 120, 240, 480, and 960 fps). An asymmetric sigmoid was fit to the VF adduction data, using five different fitting strategies with different restrictions (e.g., extended data points, fixing boundaries of the curve, etc.). The overall adduction trajectory and maximum adduction velocity were compared between fitted models and actual HSV data as a function of frame rate and fitting strategy.

RESULTS AND DISCUSSION

Regardless of frame rate, simulated data were generally well-fit by the sigmoid model. However, when compared to the actual 1000 fps data, sigmoid fits generally overestimated maximum angle velocities at lower fps. Errors in both the trajectories and maximum velocities were reduced at higher frame rates, reaching stable values by 120 fps. Noticeable differences in errors were observed by fitting strategy, with an interaction of fitting strategy and frame rate (i.e., the best fitting strategy at the lowest frame rate was not the best strategy at 120 fps and higher).

CONCLUSION

Sigmoidal models may be useful descriptors of kinematic behavior during VF adduction at frame rates of 120 fps or higher.

ACKNOWLEDGMENTS

3D-HIGH-SPEED VIDEOENDOSCOPY – PRELIMINARY FINDINGS

Michael Döllinger[1], Anne Schützenberger[1], Rita R Patel[2], Marion Semmler[1]

[1] Dep. of Otorhinolaryngology, Div. of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany
[2] Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN, USA

Keywords: 3D-high-speed videoendoscopy

INTRODUCTION

It is known that vocal folds oscillate in all three physical dimensions during phonation. However, standard imaging examinations of structural and functional characteristics in the larynx commonly only apply 2D imaging. Endoscopic laser-based 3D imaging presents the opportunity to reconstruct and analyze the 3D oscillation behavior along the entire superior vocal fold surfaces. In order to support clinical diagnostics in future, obtained 3D surface information has to be reduced to only a few significant and diagnostically conclusive parameters. However, definition of 3D amplitudes and velocities is difficult due to the absence of distinct markers comparable to the glottis contour in 2D.

METHODS

3D-High-speed videoendoscopy (HSV) was performed using a clinical PentaxMEDICAL system (512 x 256 pixel, 4000 fps) coupled to a laser projection unit, Figure 1.
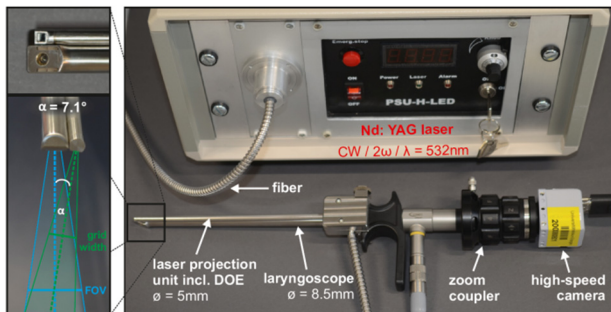
Fig. 1: Laser projection unit coupled to the endoscope.

Recordings are performed for 10 healthy subjects (5 women and 5 men). Two different approaches to calculate 3D parameters on the superior vocal folds are proposed. The resulting parameters are tested against following two hypotheses: (1) Healthy phonation is commonly associated with symmetric vocal fold oscillations. (2) Based on previously performed hemi-larynx experiments, we assume the vertical oscillation amplitudes to reach approx. 80% of the lateral amplitudes.

RESULTS

First, the measuring concept, design of the laser projection unit and safety issues will be discussed. In contrast to previous ex vivo studies, we found that in 80% of the data the maximum vertical displacement is more prominent than the maximum lateral displacement. The data suggests that healthy phonation also shows asymmetric oscillation patterns. Asymmetries in the lateral and vertical direction appear independently from each other. The dynamic vertical asymmetries seem to be more pronounced than lateral ones.

DISCUSSION & CONCLUSION

This preliminary study shows that 3D-HSV provides new information on vocal fold dynamics. However, to strengthen the presented results extensive studies are further required.

ACKNOWLEDGMENTS

AUTOMATION SUPPORT USING NON-INVASIVE MEASURES OF OPERATOR VOCALIZATION (ASIMOV)

Seth Elkin-Frankston[1], Miriam van Mersbergen[2], Bethany K. Bracken[1], Calvin Leather[1]

[1] Charles River Analytics, Cambridge, MA, U.S.A.
[2] School of Communication Sciences and Disorders, University of Memphis, Memphis, TN, USA

Keywords: **Voice control, EMG, EGG**

Abstract:

Heightened states of arousal can be investigated using electroglottography (EGG) contact quotient, as well as other measures of vocalization, such as surface laryngeal electromyography (sEMG), and accelerometry (measuring vocal cord vibrations). This type of data can be sensed using voice measures, however sophisticated conditional data processing and interpretation is required to extract these indicators. Current signal processing algorithms developed for laryngeal sEMG data are focused on clinical applications (e.g., diagnosing dysphagia) rather than detecting subtler psychological changes, and cannot readily be adapted to psychologically-based physiological changes. Novel methods are needed to clean and filter data to identify features that support assessment of arousal states. We developed and evaluated models to classify affect using a data from prior studies (van Mersbergen & Delany, 2014), consisting of EGG data, which reflects intrinsic laryngeal musculature and contains information about glottic contact area, and audio data recorded from a microphone. We chose a support vector machine (SVM) to classify data and determine affective state. Preliminary results demonstrate that our models were capable of accurately classifying negative affect. However, while negative affect could be differentiated from neutral affect, classification among positive and negative affect was less accurate. This is likely due to the fact that affect, negative or positive, are both activated through the autonomic nervous system with similar presentations. However, the goal of developing SVM models was to demonstrate our ability to successfully distinguish state. Given the accuracy of our models using limited data sets, we anticipate significant applications and performance gains when applied to new and existing voice data. Results demonstrated the utility of applying machine learning approaches to the field of voice research.

.

ACCELEROMETER-BASED AERODYNAMIC MEASURES FOR SUBJECTS WITH PHONOTRAUMATIC HYPERFUNCTION: AN IN-LABORATORY STUDY

Víctor M. Espinoza[1,5], Jarrad H. Van Stan[2,3,4], Daryush D. Mehta[2,3,4], Robert E. Hillman[2,3,4], Matías Zañartu[1]

[1] Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaiso, Chile.
[2] Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts, USA.
[3] MGH Institute of Health Professions, Massachusetts General Hospital, Boston, Massachusetts, USA.
[4] Harvard Medical School, Harvard University, Boston, Massachusetts, USA.
[5] Department of Sound, Universidad de Chile, Santiago, Chile.

Keywords: Voice Assessment; Aerodynamic Measures; Vocal Hyperfunction.

INTRODUCTION

This work aims to extend the results shown in Espinoza et al. (2017) [1], by using an Impedance-based Inverse Filtering (IBIF) algorithm [2] to estimate the glottal (airflow) volume velocity (GVV) from a neck-surface acceleration (ACC) signal. During the same experiments described in [1], a synchronized neck skin acceleration signal was recorded along with oral airflow. The ACC-based aerodynamic measures were estimated using subject-specific Q parameters from sustained vowels /a/ and /i/ in comfortable and loud loudness conditions, following the method described in [2]. For the analysis, specific voice segments (tokens) were selected in female patients with phonotraumatic hyperfunction with selected matched normal controls. Aerodynamic measures were estimated and compared with those in [1]. These initial results show that ACC-based aerodynamic measures are in close agreement with those derived from oral airflow recordings in [1].

METHODS

Estimates of peak-to-peak airflow (ACFL), maximum flow declination rate (MFDR), and open quotient (OQ) were obtained from neck-surface accelerometer recordings in 16 women with organic vocal fold lesions (nodules and polyps) and a paired control group with normal voices. In Espinoza et al. (2017) [1], glottal airflow measures were reported for selected tokens of /pae/ gestures from simultaneous and synchronous recordings of a circumferentially vented (CV) pneumotachograph mask, neck-surface acceleration, and microphone signals. To obtain an inverse filtered approximation of glottal airflow from the ACC signal, a subject-specific set of Q parameters from the IBIF algorithm was obtained. Q parameters were not derived and tested for same tokens. Instead, several sustained vocal gestures were recorded separately and used to estimate IBIF parameters. Synchronous recordings of oral airflow and neck skin acceleration signals were obtained when each subject performed a series of sustained vowels gestures (/a/ and /i/) with a constant pitch in comfortable and loud loudness condition. For each gesture, a bandpass filtered (60-1100 Hz) oral airflow vowel segment was used to perform inverse filtering with a single notch filter (SNF) constrained to a unitary gain at DC [3]. The main criterion to set SNF parameters was the minimization of formant ripple in the closed phase. Once a glottal airflow approximation is obtained from the CV mask, Q parameters were estimated using the optimization scheme depicted in [2]. A unique Q set was determined and visually confirmed (i.e., when waveforms from both oral airflow and accelerometer were reasonably similar) with a Matlab GUI designed to assess inverse filtering by a trained user. Aerodynamic measures were estimated from the inverse-filtered accelerometer signal during the same /pae/ tokens analyzed in [1] to allow for direct comparison.

RESULTS

Table 1, shows ACC-based aerodynamic measures for ACFL, MFDR, and OQ. The results from Table 1 are similar to those in Table 4 for the oral volume velocity (OVV) based measures in [1], but some differences are present, as highlighted in Table 2. To facilitate this comparison, Table 2, shows the trimmed mean (20%) differences between OVV-based and ACC-based aerodynamic measures, normalized to OVV-based measures. As is observed, MFDR has the largest differences between -2.7 and 7.4 %, followed by OQ (ranging from -5.0 to -0.5 %). ACFL is the least affected aerodynamic measure with a variation between -1.5 to 1.2%.

Table 1: Group means (standard deviation) for ACC-based aerodynamic measures for the /pae/ syllable productions in comfortable (upper row for each measure) and loud (lower values for each measure) voice for the PVH (n=16) patient group and associated matched control group with normal voices.

Measure	Control Group	PVH Group
ACFL (mL/s)	205 (71)	321 (127)
	258 (101)	422 (221)
MFDR (L/s²)	277 (105)	400 (197)
	399 (161)	657 (375)
OQ (%)	70.5 (12)	86.3 (10)
	69 (13)	82.8 (12)

Table 2: Trimmed means (20%) of the relative differences between of the OVV-based and ACC-based aerodynamic measures from the /pae/ syllable productions in comfortable (upper row for each measure) and loud (lower row for each measure) voice for the PVH patient group and associated matched control group with normal voices.

Measure	Control Group	PVH Group
ACFL (% error)	-1.5	-1.2
	-0.5	1.2
MFDR (% error)	4.1	7.4
	-2.7	7.2
OQ (% error)	-1.7	-0.5
	-5.0	-2.8

DISCUSSION

The glottal airflow measures derived from inverse-filtered neck-surface surface acceleration are in a similar range as those derived from the CV mask. Inverse filtering was challenging in this study, since female voices had high pitch and pathological conditions, that can affect both the CV and IBIF estimates [4]. However, these initial results show an overall agreement between ACC and OVV-based aerodynamic measures, indicating that these acceleration-based measures are capable of capturing vocal hyperfunctional compensatory mechanisms that have been recently reported [1]. This finding is critical for advancing toward an ambulatory aerodynamic assessment of vocal function.

CONCLUSION

Results of this study generally agree with those derived from the CV mask in Espinoza et al. (2017) [1]. The differences may be explained due to the uncertainties of both CV and IBIF-based inverse filtering methods. These results support the use of accelerometer-based IBIF measures to discriminate between vocal hyperfunction in ambulatory scenarios. Research on advanced inverse filtering methods and non-phonotraumatic vocal hyperfunction analysis is underway.

REFERENCES

[1] Espinoza VM, Zañartu M, Van Stan JH, Mehta DD, Hillman RE. Glottal Aerodynamic Measures in Women With Phonotraumatic and Nonphonotraumatic Vocal Hyperfunction. *Journal of Speech, Language, and Hearing Research* 2017; 60(8):2159-2169.

[2] Zañartu M, Ho JC, Mehta DD, Hillman RE, Wodicka GR. Subglottal impedance-based inverse filtering of voiced sounds using neck surface acceleration. *IEEE Transactions on Audio, Speech, and Language Processing* 2013; 21: 1929–1939.

[3] Perkell JS, Holmberg EB, & Hillman RE. A system for signal processing and data extraction from aerodynamic, acoustic, and electroglottographic signals in the study of voice production. *The Journal of the Acoustical Society of America* 1991; 89:1777–1781.

[4] Espinoza VM, Mehta DD, Van Stan JH, Hillman RE, Zañartu M. Uncertainty of glottal airflow estimation during continuous speech using impedance-based inverse filtering of the neck-surface acceleration signal. 173rd Meeting of the Acoustical Society of America, Boston, MA, USA. *The Journal of the Acoustical Society of America* 2017; 141(5):3579-3579.

AUDITORY-PERCEPTUAL JUDGMENTS OF LISTENER COMFORT FOR ESOPHAGEAL AND TRACHEOESOPHAGEAL SPEAKERS

Mojgan Farahani[1], Philip C. Doyle[1,2], Vijay Parsa[3,4]

[1] Graduate Program in Health & Rehabilitation Sciences, UWO, London, Ontario, Canada
[2] Voice Production & Perception Laboratory, UWO, London, Ontario, Canada
[3] School of Communication Sciences & Disorders, UWO, London, Ontario, Canada
[4] Department of Electrical & Computer Engineering, UWO, London, Ontario, Canada

***Keywords*: Listener Comfort, Tracheoesophageal Speech, Esophageal Speech**

INTRODUCTION

The consequences of total laryngectomy are significant and far-reaching. Patients will experience a variety of changes involving voice, speech, and swallowing (Blom, Singer & Hamaker, 1998), and normal speech production will be lost. These individuals are required to learn one of three "alaryngeal" modes of verbal communication. These include esophageal speech (ES) and tracheoesophageal (TE) speech (Doyle, 1994).

Regardless of alaryngeal speech production mode, there are still physical, social and psychological consequences for the laryngectomized patients (Newell, 2007). Over many decades, these alternative voice/speech modes have been studied from various aspects in order to assess performance. In the current study, the perceptual dimension of "listener comfort" (LC) for two groups of alaryngeal speakers (TE and ES) was studied in order to obtain information on whether either of these modes is perceptually judged more favorably relative to LC via a paired comparison (PC) or slider paradigm.

METHODS

Speech stimuli were obtained from six TE (3 males, 3 females) and six ES speakers (3 males, 3 females) reading the Rainbow Passage (Fairbanks, 1960); the second sentence ("The rainbow is a division of white light into many beautiful colors") was extracted and used in the auditory-perceptual phase of study. Six males and four females (n=10) served as listeners (raters) for the perceptual stimuli. All were normal hearing adults, and all were native English speakers; listeners ranged in age from 18 to 35. All were judged to be naïve listeners with no formal training and/or education in voice disorders, or laryngectomy, or the perceptual evaluation of voice. They also reported no history of speech, hearing or language difficulties. Each listener took part in two rating sessions: the first session to rate speech stimuli through a computerized paired comparison evaluation procedure which was developed in MATLAB and a slider task. The second session, for which eight of the ten listeners returned, involved rating the stimuli using the slider only. In the PC task, all 12 speaker files were played in pairs in random presentations. Each of the 12 speaker samples was paired with the other 11 speakers, and they were presented in both A-B and B-A pairings (n= 132) to determine if primacy or recency influenced listener judgments. Twenty pairs were randomly selected and presented a second time to assess intra-rater reliability. Thus, listeners rated 152 stimulus pairs. For each pair, listeners could replay the files as many times as they wished prior to identifying speaker preference. Listener comfort was defined as: "*How comfortable you would feel listening to the person's speech in a social situation*" (O'Brian et al., 2003).

The listener comfort ratings were acquired using a custom software program written in C# using Microsoft Visual Studio 2008. The graphical user interface consisted of 12 loudspeaker icons – which were randomly associated with the stimuli from the 12 speakers- and 12 associated sliders with a range between 0 – 100. Three descriptors were provided to the left of the sliders at the top, middle, and bottom: very comfortable, comfortable, and very uncomfortable. Listeners could click on the loudspeaker icons and listen to all the 12 speakers and move the sliders up and down to rate the speakers based on the LC dimension. Listeners could place the slider at any point between these headings to represent their judgment of the stimuli. Listeners were asked to listen to all the speakers first and then start ranking and rating samples based on the LC relative to other speakers.

RESULTS

Paired Comparison

Both intra- and interrater reliability measures were calculated. Intra-rater reliability was a calculation of point-by-point agreement between the original and second ratings of the 20 duplicated stimulus pairs. The reliability displayed by all the listeners was good (i.e., ≥ 0.7) with scores ranging from 0.70 – 1.00. Interrater reliability for the PC task was also calculated using Cronbach's alpha. All the listener judgments for the PC task were compared to each other and a score of 0.97 was achieved. Results indicated a high inter-rater reliability. For the PC ratings, based on the preference choices, the perceived LC rankings were calculated and a hierarchy of speakers from 1st through 12th was generated (Figure 1). Listener comfort for each speaker and each group's average were calculated

to demonstrate overall performance. The mean of the ES group was 35.61 and the TE group 61.27.

A-B & B-A stimuli positions

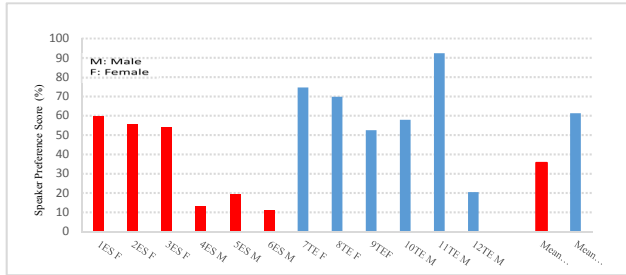The audio pairs were presented in A-B and B-A formats.

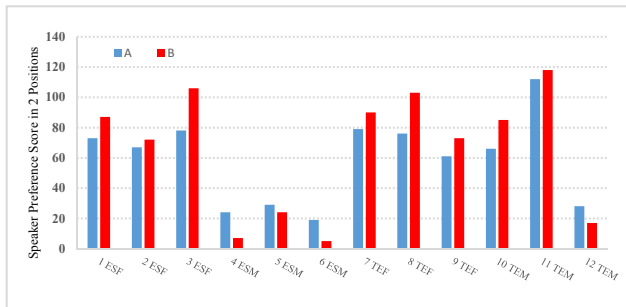**Fig. 1: PC listener comfort percentages**

Fig. 2: Stimulus presentation in AB-BA positions

Each speaker was presented once in the A position and once in B position for all pairs (Figure 2). This was done to assess if the recency or primacy of presentation order had any influence on the auditory- perceptual assessment of the stimuli. A paired-samples t-test was conducted to compare positions A and B and the possible influence of either positions on the ratings. Results showed there was no significant difference in the scores by presentation position A (M= 59.33, SD= 28.37) and position B (M= 65.58, SD= 41.01) conditions; t (11) = -1.42, p = 0.183.

Slider Task

Eight of the ten listeners completed two sets of ratings for each speaker for test-retest reliability assessments. The correlations ranged between 0.85 and 0.96 which indicated excellent reliability. Interrater reliability for the slider task was 0.98 using Cronbach's alpha. In addition, an average rating was generated for each speaker by considering all listeners' ratings of each speaker (Figure 3) and speakers were ranked. Group means were also calculated for this task (ES: 40.24; TE: 64.39) (Figure 3). A repeated measures ANOVA was conducted for statistical comparison of the effect of speaking mode (ES & TE) on speakers. The significance level was set at 0.05 for all the statistical tests. Significant effects were found for ES vs. TE group effects (F (11, 99) = 59.96, p < 0.05, ηp^2 = 0.870). In addition, post-hoc comparisons of talker scores with Bonferroni correction revealed that the mean score for the speakers: ES 4 (M = 12.3, SD = 12.27), ES 5 (M = 18.8,

SD = 11.27), ES 6 (M = 12.7, SD = 15.64), TE 11 (M = 94.9, SD = 6.66) and TE 12 (M = 22.55, SD = 12.08) were significantly different than others.
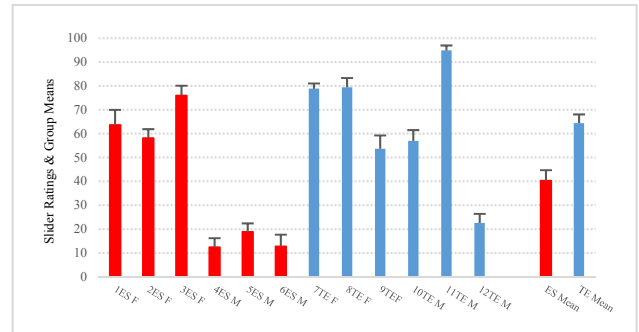
Fig. 3: Mean slider ratings and their standard errors

DISCUSSION AND CONCLUSIONS

The results from PC evaluation showed that listeners identified greater LC for the TE speakers. The slider rating tasks showed similar results and indicated that the three most preferred speakers were in the TE group. The repeated measures ANOVA indicated a significant difference in slider ratings between ES and TE speakers. Taken together, these results suggest that speech mode does have an effect on listeners' perception along the comfort dimension. Specifically, our results suggest that most of our ES speakers are rated lower on LC dimension. Overall, there was no significant difference in the scores by presentation position. These results may provide additional information to those who are deciding on a post-laryngectomy alaryngeal speech mode.

ACKNOWLEDGMENTS

REFERENCES

Blom, E.D., Singer, M.I., & Hamaker, R.C. (1998). *Tracheoesophageal voice restoration following total laryngectomy.* San Diego, CA: Singular.

Doyle, P. C. (1994). *Foundations of voice and speech rehabilitation following laryngeal cancer.* San Diego, CA: Singular.

Fairbanks, G. (1960). *Voice and articulation drillbook,* 2nd. New York: Harper & Row. pp 124-139.

Newell, C., J., S. (2007). *Gender perception for tracheoesophageal speech: a comparative evaluation of paired comparison and visual analog scaling paradigms.* (Master's thesis). Western University, London, Canada.

O'Brian, S., Packman, A., Onslow, M., Cream, A., O'Brian, N & Bastock, K. (2003). Is listener comfort a viable construct in stuttering research? Journal of Speech, Language and Hearing Research, 46 (2), 503.

EFFECT OF VOCAL TRACT CONSTRICTION ON GLOTTAL MAXIMUM FLOW DECLINATION RATE USING TOMOGRAPHIC PARTICLE IMAGE VELOCIMETRY MEASUREMENTS IN EXCISED CANINE LARYNGES

Charles Farbos de Luzan[1], Sid M. Khosla[1], Liran Oren[1], Alexandra Maddox[1], Ephraim Gutmark[2]

[1] Department of Otolaryngology-Head and Neck Surgery, Univ. of Cincinnati Medical Center, Cincinnati, Ohio
[2] Department of Aerospace Engineering and Engineering Mechanics, Univ. of Cincinnati, Cincinnati

***Keywords*: e.g. voice, glottal flow, PIV, experiments**

ABSTRACT

Objectives

Previous studies have shown that vocal intensity is highly correlated with the maximum flow declination rate (MFDR) that occurred during closing. During phonation, both parameters can be increased by increasing the subglottal glottal pressure, which in turn can reduce vocal efficiency. Voice therapies, such as the semi-occluded vocal tract (SOVT), assume that vocal efficiency can be increased by increasing the inertance (i.e. constriction) of the vocal tract. Our previous studies have shown that stronger intraglottal vortices develop near when the vocal tract constriction is increased. The aim of the current study is to show how vocal tract constriction affects the intraglottal vortices and subsequently MFDR and vocal intensity.

Methods

Flow measurements using particle image velocimetry (PIV) were taken simultaneously with acoustic measurements in excised canine larynges. vocal tract model was placed above the glottis and its constriction was varied by changing the minimal gap between the false vocal folds (FVF). Intraglottal velocity measurements were taken using 2-D PIV and volume velocity measurements were taken using tomo-PIV at the glottal exit. Testing was done at low/high subglottal pressures with low/high vocal tract constriction.

Results

Figure 1 shows and example of the time-resolved data obtained from tomo-PIV. The time-averaged volume flow rate compares well with the reference value given by a Coriolis flow-meter (Fig. 2). The location of the maximum flow declination rate is indicated by a yellow cross, and occurs at the beginning of closing.
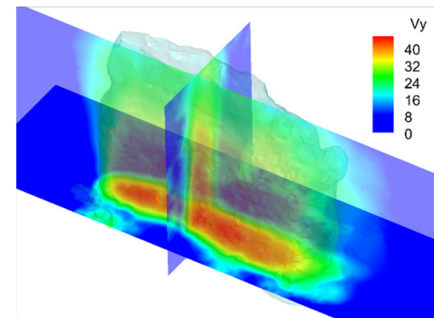
Figure 1 The phase averaged flow fields allow to clearly visualize the 3D glottal jet. Plots show contours of velocity magnitude along the mid-coronal plane, the transverse plan directly above the glottal exit and a 3D iso-surface of velocity.
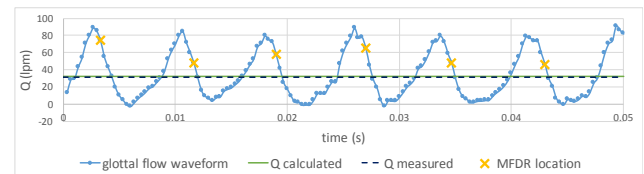
Figure 2 An example of the computed glottal flow waveform from PIV measurements at 18 cmH2O.

For a given subglottal pressure, MFDR, the acoustic intensity (SPL), and vocal efficiency all increased as the constriction between the FVF was increased. The vortical strength, MFDR, and SPL also increased as a function of the subglottal pressure when the vocal tract was removed.

Conclusions

Changing the strength of the intraglottal vortices can affect MFDR and subsequently vocal efficiency. Therefore, vocal efficiency can be increased by increasing vocal tract inertance (as in SOVT) or by increasing the strength of the vortices. Mechanisms for the latter and their clinical implications will be further discussed.

GLOBAL AND LOCAL DIFFERENTIATION OF HUMAN VOCAL FOLD FIBROBLAST GENOTYPE

Alexander G. Foote, M.A., CCC-SLP[1], Ziyue Wang[2], Christina Kendziorski, PhD[3], Susan L. Thibeault, PhD, CCC-SLP[4]

[1,4] Division of Otolaryngology, Department of Surgery, University of Wisconsin-Madison, Madison, Wisconsin, USA
[2] Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin, USA
[3] Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Keywords: **Vocal fold biology; human vocal fold fibroblast; functional gene expression; transcriptome profiling, mechanobiology**

OBJECTIVES

At present, many physiological and pathophysiological aspects of vocal fold fibroblast (VFF) function remain poorly understood, however it is becoming established that physical forces, such as mechanical stress, are essential for tissue homeostasis and may influence gene expression of cells. Significant challenges exist with regard to tissue engineering of the vocal fold lamina propria after wound injury and its unique biomechanical characteristics have not yet been entirely understood. The purpose of this study was to compare the VFF genotype with their cellular counterparts at various anatomic locales to identify differences in functional gene expression profiles.

METHODS

In vitro. We investigated 7 normal human fibroblast primary cell lines from healthy cadavers, which included: vocal fold, trachea, lung, abdomen, scalp, upper gingiva, and soft palate. Primary cultures were validated with a subtractive immunocytochemical methodology for confirmation of fibroblast lineage and next generation RNA-sequencing was performed for gene expression analyses.

RESULTS

Unsupervised gene expression analysis yielded 3999 genes differentially expressed with the false discovery rate controlled at 5%. Hierarchical cluster analysis of the data revealed fibroblast grouping based on anatomic site origin rather than cells from the same donor, suggesting global phenotype heterogeneity. Sex and age-related effects were negligible. Functional enrichment analyses revealed several functional themes related to transcription factors for signaling pathways regulating pluripotency of stem cells, as well as, extracellular matrix components such as cell signaling, migration, proliferation, and differentiation potential.

CONCLUSIONS

Human fibroblasts display a phenomenon of global topographic differentiation, which is maintained in isolation via *in vitro* assays. Epigenetic mechanical influences on vocal fold tissue may play a role in uniquely modelling and maintaining the local environmental cellular niche during homeostasis with VFF distinctly specialized related to their anatomic positional and developmental origins established during embryogenesis.

ACKNOWLEDGEMENTS

EFFECTS OF THE VOCAL TRACT ON PHONATION THRESHOLD PRESSURE

L. Fulcher[1], A. Lodermeyer[2], G. Kähler[2], M. Döllinger[3], S. Becker[2], and S. Kniesburges[3]

[1] Department of Physics & Astronomy, Bowling Green State University, Ohio, USA
[2] Process Machinery & Systems Engineering, Friedrich-Alexander University, Erlangen, Germany
[3] Phoniatrics & Pediatric Audiology, University Hospital, Erlangen, Germany

Keywords: Vocal Tract, Synthetic Vocal Folds, Phonation Threshold Pressure, Surface Wave Model

INTRODUCTION

Accelerating the air within the vocal tract has an effect on the pressure of the air flowing through the glottal channel. This effect may be described by an equation involving the product of the vocal tract inertance I_{VT} and the rate of change of the volume velocity U_{VT} within the vocal tract [1-3], that is,

$$P = I_{VT}\, dU_{VT}/dt, \qquad (1)$$

The inertance of the vocal tract is given by the espression, $I_{VT} = \rho\, L_{VT}/A_{VT}$, where ρ is the density of the *air*, L_{VT} is the length of the vocal tract, and A_{VT} is its cross-sectional area. If (1) is combined with the dynamics of a lumped element model, such as Titze's surface wave model (SWM), then the effects of adding a vocal tract to a synthetic vocal fold model without a vocal tract are completely determinded, since (1) contains no free parameters. The purpose of this study is to examine the validity of (1) for describing the pressure changes due to a vocal tract by comparing calculations based on it with measurements taken with vocal tracts of different cross sectional areas (2.7 cm^2 to 11.7 cm^2) and of different lengths (4.0 cm to 59.0 cm).

METHODS

In the SWM the equation of motion for the vocal fold coordinate ξ is

$$m\,\ddot{\xi} + B\,\dot{\xi} + K(\xi)\,\xi = L_g T\, P_g(t) = \int_{-T/2}^{T/2} P(z)dz, \quad (2)$$

where m is the mass of the vocal fold, B is the damping parameter, $K(\xi) = k\,(1 + \eta\,\xi^2)$ is the nonlinear stiffness, T is the glottal thickness, and L_g is the glottal length. The quantity $P_g(t)$ is the average of the pressure over the medial surface of the vocal fold. Assuming a trapezoidal shape for the prephonatory profile of the vocal fold, the integral in (2) can be done analytically [3], which yields

$$P_g(t) = P_{sub}\, \frac{1 - A_2(t)/A_1(t)}{1 + (k_{ent}-1)A_2^2(t)/\, A_1^2(t)}, \qquad (3)$$

where $A_1(t)$ and $A_2(t)$ are the areas of the glottal entrance and the glottal exit, and k_{ent} is the entrance loss coefficient. An essential aspect of the SWM is that these two areas may be expressed in terms of the vocal fold coordinate and its velocity so that the ratio of Eq. (3) takes the form,

$$\frac{A_2(t)}{A_1(t)} = \frac{\xi_0 + \xi(t) - \tau\dot{\xi}(t)}{\xi_0 + \xi(t) + \tau\dot{\xi}(t)}, \qquad (4)$$

where τ is the time for the surface wave to propagate from the center of the medial surface of the vocal fold to the top, and ξ_0 is the prephonatory glottal half width. The quantity $\tau = T/\ (2\ c)$, where c is the wave velocity. Solving (2) – (4) numerically yields results to compare with the the data taken by Chan, Titze, and Titze without a vocal tract [4], as shown in Fig. 1. The parameters E and F in Fig. 1 are used to determine the entrance loss coefficient k_{ent}. The nonlinear parameter η of Eq. (2) is set equal to zero.
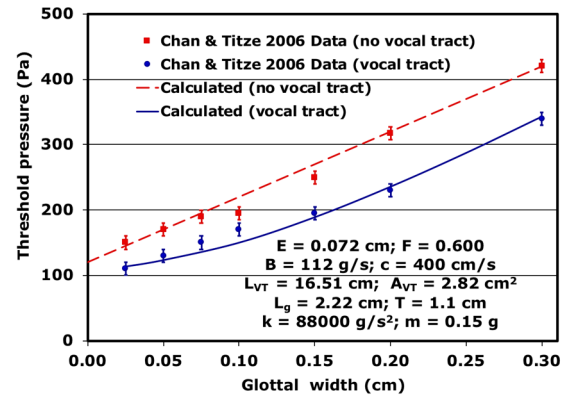
Fig. 1: Measured and calculated threshold pressures as functions of glottal halfwidth with and without a vocal tract.

The effects of the vocal tract may be included by adding (1) to the equation that determines the volume velocity. Since the length of the vocal tract (16.51 cm) and the area of the vocal tract (2.82 cm^2) are given by the geometry of the experiment, adding the vocal tract does not introduce additional free parameters into the calculation.

RESULTS AND DISCUSSION

The vocal tract with an adjustable lateral dimension D is shown in Fig. 2. The parameter D can be adjusted from 1.8 cm to 7.8 cm. Since the perpendicular dimension of the vocal tract is 1.5 cm, the area of the vocal tract ranges from 2.7 cm^2 to 11.7 cm^2.
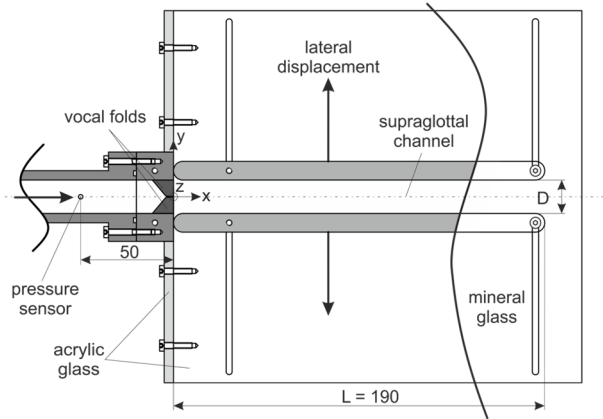
Fig. 2: Schematic diagram of the supraglottal channel with an adjustable lateral dimension D. All distances are in millimeters.
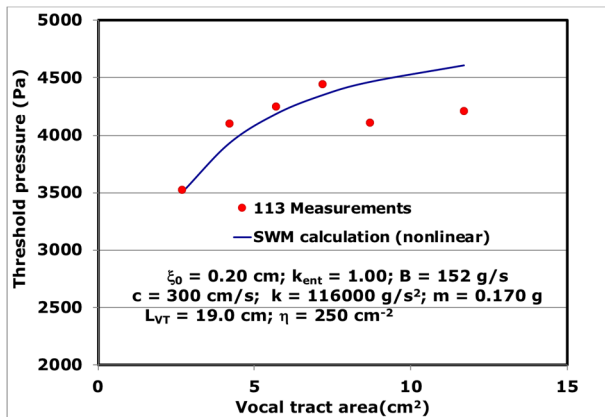
Fig. 3: Measured and calculated threshold pressures taken with vocal fold replica 113 for different vocal tract areas.

Fig. 4: Measured and calculated frequencies at threshold taken with vocal fold replica 113 for different vocal tract areas.
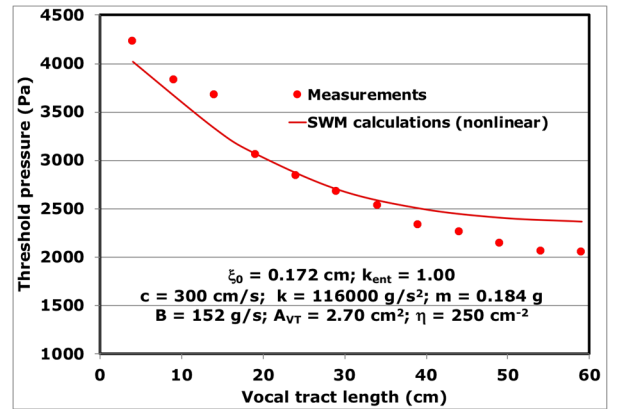
Fig. 5: Measured and calculated threshold pressures taken with vocal fold replica 113 for different vocal tract lengths.

Measurements of threshold pressures for different vocal tract areas are shown in Fig. 3. These are compared with numerical calculations based on (1) – (4) above. Agreement is reasonable for the 4 smaller areas, but the discrepancy is considerable for the larger areas. Although we are not certain of the source of the discrepancy, it may be due to a change in the flow pattern near the glottal exit. The label 113 describes the mixture of Ecoflex 30 and Silicone Thinner used to construct the vocal fold replica. Measured frequencies at threshold for the same vocal fold replica are shown in Fig. 4. The nonlinear parameter η is essential for the good agreement of the calculations and measurements shown in Fig. 4.

Measurements of the threshold pressure with vocal tracts of different lengths are shown in Fig. 5. The elastic parameters there were taken to be the same as those of Figs. 3 and 4, since the 113 replica was also used for these experiments. Agreement with the calculations is very reasonable. These data required the construction of a vocal tract section 4 cm long and 11 sections 5 cm long.

CONCLUSIONS

Equation (1) states that the effect of the vocal tract on properties of phonation near threshold involve its inertance, with a direct dependence on its length and an inverse dependence on its cross-sectional area. Our calculations with the SWM model and the measurements described here support this claim.

REFERENCES

[1] Ishizaka K, Flanagan J. Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Sys. Tech. J 1972; 52:1233-1268.

[2] Titze I. The physics of small amplitude oscillation. J. Acoust. Soc. Am. 1988; 83:1536-1552.

[3] Fulcher L, Scherer R. Phonation threshold pressure: Comparison of calculations and measurements taken with physical models of the vocal fold mucosa. J. Acoust. Soc. Am. 2011; 130:1597-1605.

[4] Chan R, Titze I, and Titze M. Further studies of pnonation threshold pressure in a physical model of the vocal fold mucosa. J. Acoust. Soc. Am. 1997; 101: 3722-3727.

A PRELIMINARY COMPARISON OF THE EXPRESSION AND LOCALIZATION OF INTEGRAL EPITHELIAL PROTEINS IN HUMAN AND RABBIT VOCAL FOLDS

Gary Gartling, MS[1]; Lea Sayce, DPhil[2]; Emily Kimball, MS[1]; Shintaro Sueyoshi, MD[2];

Jennifer Brandley, BA[1]; Bernard Rousseau, PhD, CCC-SLP, MMHC [123]

[1] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA.
[2] Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA.
[3] Department of Mechanical Engineering, Vanderbilt University Medical Center, Nashville, TN, USA.

Key words: Vocal fold; Epithelium; Rabbit; Human

INTRODUCTION

The presence of integral membrane transport proteins such as Sodium/PotassiumATPase and Aquaporin 1, as well as the cell adhesion transmembrane protein E-Cadherin are believed to influence vocal fold health and voice production [1-2]. Sodium/PotassiumATPase and Aquaporin 1 are thought to regulate and maintain vocal fold surface hydration, while E-Cadherin has been observed to support epithelial barrier integrity [3-6]. The expression and localization of these proteins have been described in the healthy vocal folds of canine, ovine, and porcine models [1-9]. However, despite the growing body of work utilizing New Zealand white rabbits (*Oryctolagus cuniculus*) as a model for vocal fold disorders, the expression and localization of these proteins in this species has yet to be investigated. Additionally, translatability from this animal model to humans has yet to be validated. The purpose of this preliminary study is to compare the expression and localization of these proteins between these species, and to better understand how studies in a rabbit model can inform clinical practice.

METHODS

Sample Collection

This study was approved by Vanderbilt Institutional Review Board, #170371. A donor larynx from a 44-year-old African American male was procured by pathologists from the Cooperative Human Tissue Network. The sample was harvested within 24 hours of the donor's death and transported to Vanderbilt University Medical Center, where it was cryopreserved upon arrival. The vocal fold was determined to be healthy and absent of any gross tissue abnormalities by an Otolaryngologist. For comparison, a larynx from an adult male New Zealand white rabbit was harvested using methods approved by the Institutional Animal Care and Use Committee at Vanderbilt University. Both human and rabbit larynges were hemisected, and the left half of each larynx was mounted in Optimal Cutting Temperature (OCT).

Instrumentation and Measurement

OCT blocks were sectioned in the coronal orientation at 12µm thickness using a Leica CM1900 cryostat and mounted onto positively charged slides. Samples were immunolabeled using primary antibodies against Aquaporin 1 (Santa Cruz sc-25387), E-Cadherin (BD Transduction 610182), and Sodium/PotassiumATPase (Abcam ab7671). All primary antibodies were co-labeled with a fluorescent secondary antibody, and cell nuclei were counterstained blue using DAPI. Images were captured digitally at 40x magnification using a Nikon Eclipse 90i microscope and Hamamatsu C10600 camera.

Analysis

During image capture, exposure times were kept consistent across samples. Images from each sample included the epithelium and superficial layer of the lamina propria of the sectioned vocal fold. The expression and localization of each protein of interest was qualitatively analyzed and compared between species.

RESULTS AND DISCUSSION

Similar localization of the ion transport channels Aquaporin 1 and Sodium/PotassiumATPase, and the tight-junctional protein E-Cadherin was observed between rabbit and human vocal fold tissue (Fig. 1). Aquaporin 1, a channel which passively transports water and solutes across the cell membrane, was localized to the basement membrane region of the epithelium and throughout the lamina propria. Sodium/PotassiumATPase is an ion channel which contributes to the difference in electrical charge between the interior and the exterior of the cell, thus driving trans-epithelial water transport. This channel was primarily localized to the cell membrane of basal epithelial cells. E-Cadherin, a protein involved in cell-cell adhesion and the maintenance of the epithelial barrier, was expressed at the cell junctions throughout the stratified epithelial layer.

While the study presented demonstrates similar expression and localization, the limited sample size

necessitates further investigation to verify these findings. Similarly, exploration of other key proteins within the vocal fold epithelium, lamina propria, and thyroarytenoid muscle is essential for the validation of the rabbit as a scientific model for the study of voice injury in humans.

Fig. 1: (a-c) Representative immunofluorescence images of human vocal fold (top panels) and rabbit vocal fold (bottom panels) tissue stained for DAPI (blue) and (**a**) Aquaporin 1, (**b**) E-Cadherin, and (**c**) Sodium/PotassiumATPase (green). Scale bars: 10 μm**.**

CONCLUSION

These data identify that key ion transport channels and cell adhesion proteins are similarly localized and expressed in the rabbit and human vocal folds. Additional studies and procedures are required to further compare vocal fold physiology between species and confirm these preliminary findings. However, the commonalities presently described suggest similarity of critical vocal fold components and support the use of the rabbit as a model of human vocal fold physiology and voice disorders.

REFERENCES

[1] Leydon C, Sivasankar M, Falciglia, DL, Atkins C, Fisher KV. Vocal fold surface hydration: A review. *Journal of Voice* 2009;23(6): 658-665.

[2] Erickson EL, Leydon C, Thibeault SL. Vocal fold epithelial barrier in health and injury: A research review. *Journal of Speech, Language, and Hearing Research* 2014; 57:1679–1691.

[3] Gill GA, Buda A, Moorghen M, Dettmar PW, Pignatelli M. Characterisation of adherens and tight junctional molecules in normal animal larynx; determining a suitable model for studying molecular abnormalities in human laryngopharyngeal reflux. *Journal of clinical pathology.* 2005;58(12):1265-70.

[4] Fisher KV, Telser A, Phillips JE, Yeates DB. Regulation of vocal fold transepithelial water flux. *J Appl Physiol* 2001;91:1401–1411.

[5] Lodewyck D, Menco B, Fisher K. Immunolocalization of aquaporins in vocal fold epithelia. *Archives of Otolaryngology - Head and Neck Surgery* 2007; *133*(6):557–563.

[6] Ahmed MER, Bando H, Hirota R, Sakaguchi H, Koike S, El-Adawy AASN, ... Hisa Y. Localization and regulation of aquaporins in the murine larynx. *Acta oto-laryngologica* 2012;132(4):439-446.

[7] Alipour F, Jaiswal S, Vigmostad S. Vocal fold elasticity in the pig, sheep, and cow larynges. *Journal of Voice* 201;25(2):130-136.

[8] Kurita S, Nagata K, Hirano M. Comparative histology of mammalian vocal folds. In: Kirchner JA, ed. *Vocal fold histopathology. A symposium*. San Diego, CA: College Hill Press, 1986: 1–10.

[9] Garrett CG, Coleman JR, Reinisch L. Comparative histology and vibration of the vocal folds: implications for experimental studies in microlaryngeal surgery. *The Laryngoscope* 2000;110(5):814-824.
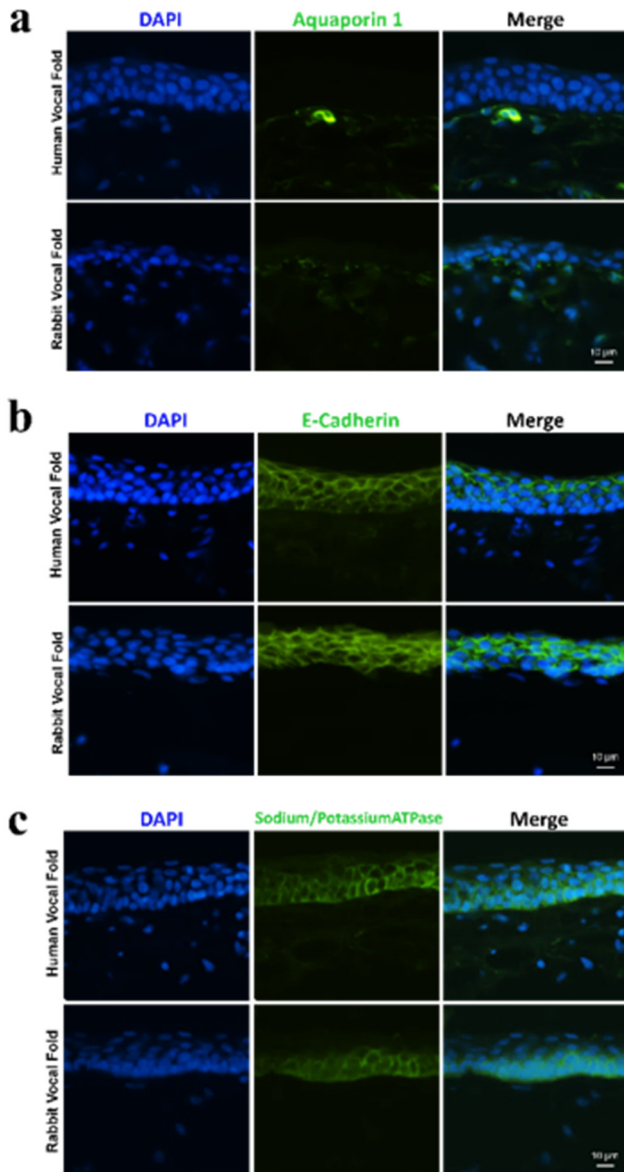
MODELING DYNAMICS OF CONNECTED SPEECH IN TIME AND FREQUENCY DOMAINS WITH APPLICATION TO ALS

Hamzeh Ghasemzadeh [1], Jeff Searl [1]

[1] Department of Communicative Sciences and Disorders, Michigan State University, East Lansing, Michigan, United States

Keywords: acoustics; connected speech; dynamics of speech; machine learning

INTRODUCTION

Using the acoustic signal during communication to gain insight into health and functioning within the body has been of interest for years. It is relatively inexpensive, easy to obtain acoustic recordings, and importantly, the collection procedure is not invasive leaving the underlying production of speech undisturbed. There have been two broad categories of acoustic samples considered for analysis over the years: sustained vowels and connected speech (words, phrases, sentences). Previous studies have argued that speech samples are more ecologically valid, more informative, and sometimes the more likely context for manifestation of certain disorders [1, 2, 3].

The speech signal is the outcome of dynamic interaction between the voice source and vocal tract filter. In connected speech there are time and frequency varying characteristics in the acoustic signal, determined by changes in the source and filter. The objective of this study was to propose and pilot test a framework for modeling the dynamics of the acoustic signal that accounts for time and frequency domains simultaneously. Our hypothesis was that the proposed model will differentiate connected speech produced by normal controls from individuals with amyotrophic lateral sclerosis (ALS) who have varying degrees of speech impairment.

METHODS

Human Data

Inclusion criteria for the ALS participants were: (a) firm diagnosis of ALS from the treating neurologist, (b) ≥ 40 years of age, (c) self-reported functional hearing in a quiet room, (d) presence of bulbar symptoms based on clinical notes of speech language pathologist (SLP), and (e) use of verbal communication for at least 25% of communication needs. Participants were excluded if they were ventilator dependent, had a positive history of stroke, head injury, or other neurological conditions besides ALS, or had a history of head and neck related surgeries that impacted speech.

Based on these criteria, recordings from 8 individuals (4 males, 4 females; age 44-77 years) over three sessions (baseline, 3 months later, 6 months later) were used for this study. For other statistics regarding this population, see [4]. Additionally, a group of 12 healthy individuals (6 male, 6 female; age 47-78 years) were recorded in two sessions (baseline, 6 months later) as a normal control group.

As part of a larger protocol [4] acoustic recordings were obtained of all participants completing the Word-In-Phrase portion of the Speech Intelligibility Test for Windows© [5]. This consisted of 57 randomly selected words inserted in the phrase "say [target words] again". Acoustic signals were recorded in a single-wall sound booth using a head mounted microphone (AKG C410) positioned 6 cm from the corner of the mouth at a 45° azimuth and a digital recorder (Zoom H4N, 24-bit/96kHz). Power of signals was normalized in Matlab to remove effect of variations in the power across participants.

Analysis

Visual representation of the acoustic signal can be a rich source of information [6]. To that end, the acoustic signal is converted into a two-dimensional signal, where one axis represents variations in the time domain and the other axis represents variations in the frequency domain. A combination of the spectrogram and quantization is one approach for such a conversion [6]. Another possible approach uses a triangular filter bank in a suitable frequency scale [7] followed by quantization.

The method proposed here is executed in three steps. First, a set of signal processing techniques are applied on the acoustic signal converting it into an image with suitable properties. Second, statistical models of the image along the x and y axes are constructed. Third, pattern recognition techniques and a machine learning algorithm are used to find discriminative patterns between the dynamics of speech from controls and participants with ALS. Figure 1(A) shows a block diagram of the approach. The results of applying the signal processing and statistical modeling on a normal control sample is depicted in Figure 1(B).

For the second step, neighboring pixels along x directions were extracted and their joint distribution was calculated. The results are an estimation of the statistical model of variations in the acoustic signal in the time domain. The same procedure was followed in the y direction of the image for estimating the statistical model of variations of the signal in the frequency domain. Machine learning in step 3 utilized a support vector machine (SVM) with radial basis function (RBF). This technique allows the system to learn complex and non-linear relationships between features (dependent variables) in high dimensional space. Also, generalization of the system was tested with 10-fold cross validation technique.
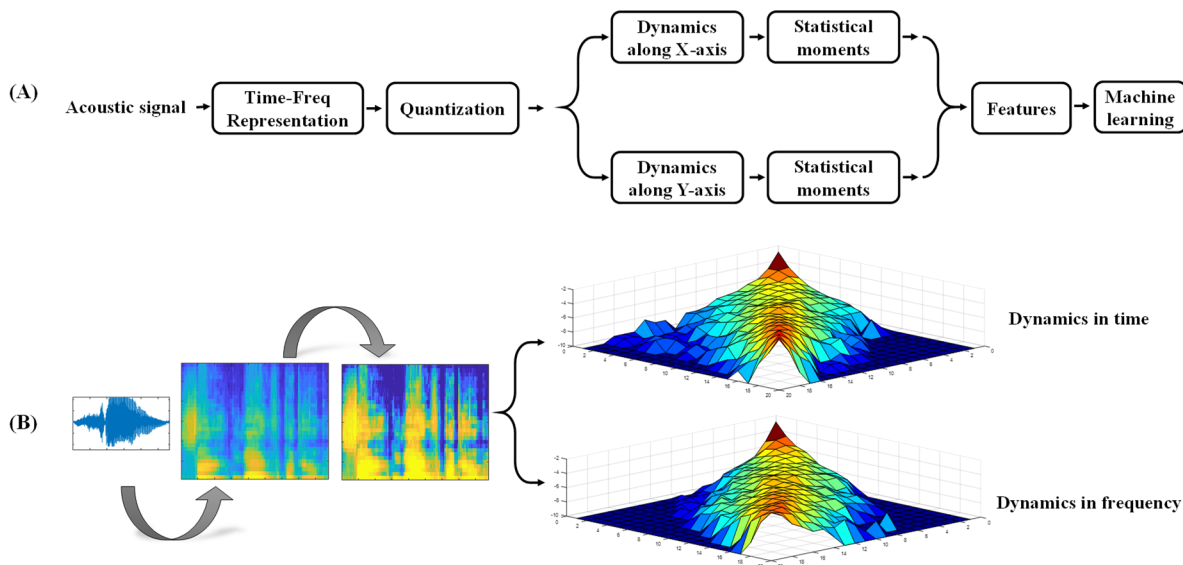
Fig. 1: (A) block diagram of the system   (B) results of applying signal processing and statistical modeling steps on a normal control sample. The three-dimension plots are a joint distribution of pixels along x and y axes of the image

RESULTS

The system was tested for dynamics in the time and frequency domains separately and the results are reported in Table 1. Sensitivity and specificity (and resulting accuracy) were high in both domains.

Table 1: performance of the system

	Time dynamics	Frequency dynamics
Sensitivity	92.5%	92.6%
Specificity	93.7%	94.2%
Accuracy	93.2%	93.4%

CONCLUSION

The proposed method has the potential of capturing dynamics of connected speech in a meaningful manner allowing separation of normal from ALS-disordered speech. The ALS samples ranged from minimally to severely impaired. Of interest is exploration of the model's ability to distinguish the most subtle speech changes in ALS, perhaps preceding the ability to detect changes perceptually, to serve as an early indicator of bulbar disease onset. The value of early detection is high particularly for pharmaceutical and other intervention studies to serve in classifying participants as having or not yet having ALS onset in the head and neck, or as a sensitive measure of change over time or drug/therapy status.

ACKNOWLEDGMENTS

REFERENCES

[1] S. N. Awan, N. Roy, and C. Dromey, "Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model," *Clin. Linguist. Phon.*, vol. 23, no. 11, pp. 825–841, 2009.

[2] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, "Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels," *J. voice*, vol. 24, no. 5, pp. 540–555, 2010.

[3] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice: sustained vowels versus continuous speech," *J. Speech, Lang. Hear. Res.*, vol. 44, no. 2, pp. 327–339, 2001.

[4] J. Searl, S. Knollhoff, and R. J. Barohn, "Lingual – Alveolar Contact Pressure During Speech in Amyotrophic Lateral Sclerosis: Preliminary Findings," *J. Speech, Lang. Hear. Res.*, vol. 60, no. 4, pp. 810–825, 2017.

[5] K. M. Yorkston, D. R. Beukelman, M. Hakel, and M. Dorsey, "Speech intelligibility test for windows," *Lincoln, NE Commun. Disord. Softw.*, 1996.

[6] H. Ghasemzadeh, M. Tajik Khass, and H. Mehrara, "Cipher text only attack on speech time scrambling systems using correction of audio spectrogram," *ISC Int. J. Inf. Secur.*, vol. 9, no. 2, pp. 33–47, 2017.

[7] H. Ghasemzadeh, M. Tajik Khass, and M. Khalil Arjmandi, "Audio steganalysis based on reversed psychoacoustic model of human hearing," *Digit. Signal Process. A Rev. J.*, vol. 51, pp. 133–141, 2016.

DIFFERENTIATION CAPABILITY OF PHONATION BIOMECHANICAL FEATURES BETWEEN PRESBYPHONIC AND PARKINSON PATIENTS' VOICE

Andrés Gómez[1], Pedro Gómez[1], Jiri Mekyska[2], Irena Rektorova[3], Agustín Álvarez[1], Victoria Rodellar[1]

[1] NeuVox Lab, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain
[2] Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
[3] First Department of Neurology, St.Anne's University Hospital, Brno, Czech Republic

Keywords: **Phonation Instability; Neuromotor Diseases; Mahalanobis Distance**

INTRODUCTION

Hypokinetic dysarthria (HD), frequently diagnosed in Parkinson's Disease (PD), and presbyphonia share common manifestations which may hamper pathology detection and monitoring based on acoustic analysis. The present work seeks to differentiate phonation feature sets specific of presbyphonic voice from those specific of HD.

METHODS

Vocal fold biomechanics, glottal closure defects and tremor are obtained by inverse filtering. Phonations from two databases are used in three-band tests. A first database (NR: normative reference) contains phonations from normophonic subjects (adult voice) in the range between 18 to 50 years-old (50 males and 50 females). A second database contains phonations from a set of 60 male and 40 female PD patients in a range of 66.3±8.6 and 69±7.7 years-old (respectively). A third database (HC: healthy controls) contains speech from a control set of 26 male and 27 female subjects in a range of 65.6±8.9 and 61.8±9.1 years old (respectively). PC and HC subjects were sampled from the PARCZ database [1].

Features from phonations of [a:] from the three databases are compared. The features used in the study are distortion (D: jitter and shimmer), biomechanical (B: body and cover mass and stiffness unbalance between neighbor phonation cycles) and neurologic (N: tremor in the bands 2-5, 5-8 and 8-11 Hz respectively). Segments lasting 300-500 ms sampled at 16 kHz were processed to remove the vocal tract resonances by adaptive LPC filtering [2]. Estimates of glottal source were used to evaluate the above described set of features. Fig. 1 shows the features of each subject selected from the three data subsets (NRS, PDS and HCS respectively). Each feature has been normalized by its corresponding average value estimated from the normative dataset (NR).

Statistical Analysis

Gender-selective evaluations based on Log-Likelihood tests [3] have been used to compare control against reference sets, pathologic against control, and pathologic against reference sets. Fig. 2 (left) shows the scatter plots of the three subsets with respect to the three most-resolving features, these being shimmer, tremor in the band 8-11 Hz and the rms amplitude of tremor.
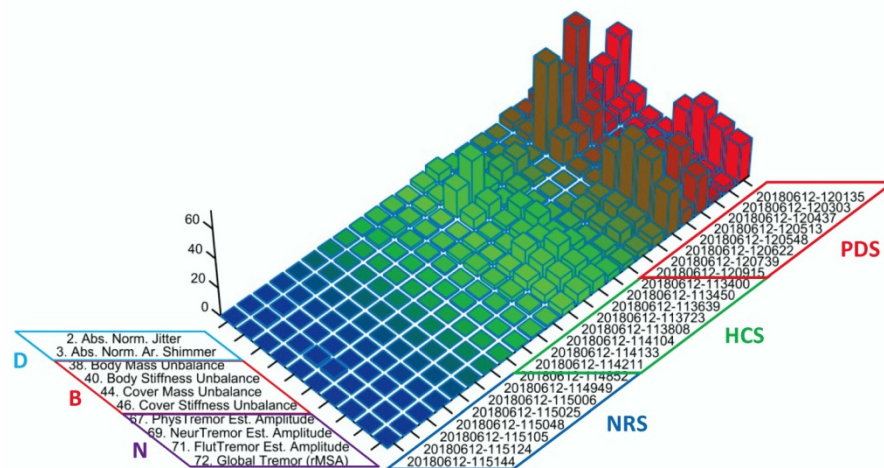
Fig. 1 : Comparison of relevant features from the three datasets normalized to their respective average values. D, B, F: distortion, biomechanical and neuromotor features; NRS, HCS, PDS: normative reference, healthy controls and Parkinson's Disease subsets.

The feature/subject plot in Fig. 1 shows that six of the PDS subjects show large distortions in most of the features

considered (except in cover unbalances), whereas two subjects presented moderate distortion. The situation in

reference to HCS subjects is more uniform, where moderate distortion may be found in most of the features considered. On its turn, NRS subjects show very low distortion. This situation is illustrated when the three subsets are clustered as shown in Fig. 2. It may be seen that the NRS (in green bullets) is tightly packed near the origin.

Tremor minimum, jitter is between 0.5 and 1.5%. HCS (blue diamonds) is more spread, showing a rectangular shape around its centroid (larger diamond). PDS (red squares) is quite widespread, two subjects closer to origin, three others being closer to centroid (larger square), and the rest quite far from the origin.
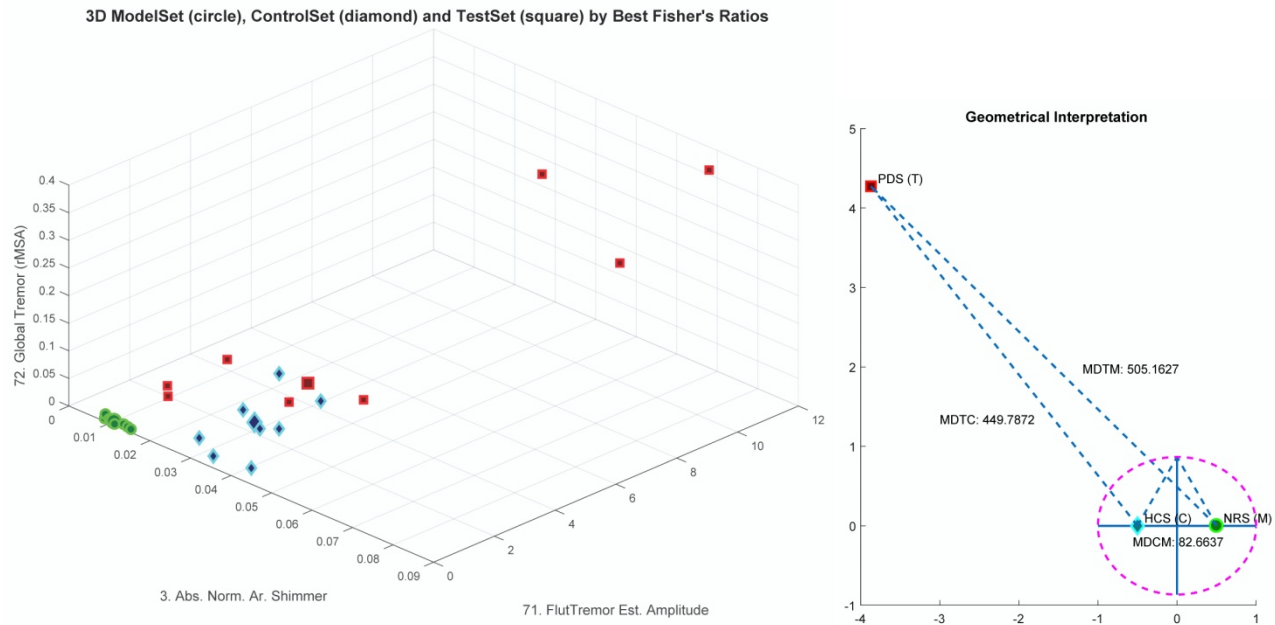
Fig. 2 : Classification results: male subset. Left: 3D scatter plot of the three subsets in the experiment in terms of the three most relevant features from their Fisher Ratios; b: Dataset centroids projected on their reference plane.

RESULTS AND DISCUSSION

Significant differences have been found between control and reference sets in the subspaces of biomechanical and contact defects, as well as in tremor features. Relevant differences between pathologic and reference sets have been found in jitter, shimmer, body feature unbalance, and tremor features. These are also present between pathologic and control sets, although some PD subjects may show moderate differences with respect to the reference set, even smaller than healthy controls. A global separation among the three clusters is shown in Fig. 2 (right), where the three cluster centroids have been plotted as in a 2D projection, based on their relative Mahalanobis distances [4], where MDCM is the distance between HCS and NRS, MDTM is the distance between PDS and NRS, and MDTC is the distance between PDS and HCR. When all the eight features are considered, it may be seen that the distance between controls and references are much smaller than the distance between pathological and references or controls. This observation points out to the need to unbias control estimates with respect to normative references when using them in comparative tests or clustering protocols.

CONCLUSIONS

The results point out to a common pattern of manifestation of HD and presbyphonia in biomechanical

features, although tremor and distortion features show statistical relevance in differentiating between both sets.

REFERENCES

[1] Mekyska J, Janusova E, Gómez P, Smekal Z, Rektorova I, Eliasova I, Kostalova M, Mrackova M, Alonso JB, Faúndez M, López de Ipiña K. Robust and complex approach of patohogical speech signal analysis, *Neurocomputing*, 2015; 167: 94-111.

[2] Deller JR, Proakis JG, Hansen JHL, *Discrete-Time Processing of Speech Signals*, NewYork, Macmillan; 1993.

[3] Gómez P, Rodellar V, Nieto V, Muñoz, C, Mazaira LM, Martínez R, Álvarez A, Ramírez C, Fernández M. Characterizing Neurological Disease from Voice Quality Biomechanical Analysis. *Cognitive Computation* 2013; 5(4):399-425.

[4] Xiang S, Nie F, Zhang C. Learning a Mahalanobis distance metric for data clustering and classification. *Pattern Recognition* 2008; 41: 3600-3612.

DIAGNOSTIC VALUE OF ARTICULATION NEUROMOTOR INSTABILITY FOR NEURODEGENERATIVE DISEASE DETECTION AND RATING

Andrés Gómez[1], Pedro Gómez[1], Jiri Mekyska[2], Irena Rektorova[3], Victoria Rodellar[1], Agustín Álvarez[1]

[1] NeuVox Lab, Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain
[2] Department of Telecommunications, Brno University of Technology, Brno, Czech Republic
[3] First Department of Neurology, St.Anne's University Hospital, Brno, Czech Republic

***Keywords*: Articulation instability; Neuromotor Diseases; Mutual Information**

INTRODUCTION

Neurodegenerative diseases leave an imprint on both phonation and articulation stability. Hypotonia (muscle stiffness), vocal fold unbalance (bi-laterality) and tremor (altered neuromotor feedback) some ways in which the degeneration manifests. Articulatory instability has been less studied, compared to reduced vowel space and vowel centralization. Another relevant aspect is the dynamic behavior of formant transitions in dyadochokinetic exercises, to assess hypokinetic dysarthria in Parkinson's Disease (PD). The objective of the study is to compare articulation in PD patients against a normative population from kinematic features estimated on formants using Information Theory to quantify their divergence with respect to normative subjects.

MATERIALS AND METHODS

Jaw-Tongue Biomechanical Model

Speech articulation depends on the activity of vocal tract structures, such as the jaw, tongue, lips and velo-pharynx, among others. These structures are moved by different muscles, which are activated by neuromotor pathways from cranial nerves [1]. The production of speech sounds depends on the positions of these structures and on their dynamic displacements.
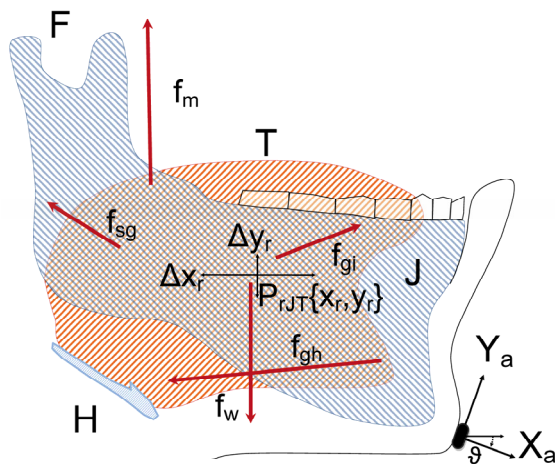
Fig. 1 : Jaw-Tongue Model. F: Fulcrum; T: Tongue; J: Jaw bone; H: Hyoid bone; f_{sg}: stylo-glossus force; f_m: masseter

force; f_{gi}: glosso-intrinsic forces; f_{gh}: genio-hyoid force; f_w: gravity; X_a, Y_a: accelerometer normal and tangential; Δx_r, Δy_r: horizontal and vertical displacements of the reference point (P_{rJT}) in the sagittal plane.

In the present paper the role of the jaw-tongue system, as depicted in Fig. 1 will be studied when affected by neuromotor degeneration induced by PD. The jaw-tongue biomechanical system is considered to be a third-order lever with lumped mass load concentrated in the reference point $\{x_r, y_r\}$. Harmonic oscillation $\{\Delta x_r, \Delta y_r\}$ around the fulcrum (F: attachment to the skull) is assumed under forces acting on this system (see Fig. 1).

A kinematic correlate of articulation

The displacement of the reference point of the jaw-tongue system when observed over time may be described by an absolute kinematic velocity (AKV) defined as:

$$|v_r| = \left[B_1 \left(\frac{dF_1}{dt}\right)^2 + B_2 \left(\frac{dF_2}{dt}\right)^2 + B_{12} \frac{dF_1}{dt}\frac{dF_2}{dt}\right]^{\frac{1}{2}} \quad (1)$$

Where F_1 and F_2 are the first two formants, and B_1, B_2 and B_{12} are quadratic scaling factors relating movement and acoustics [2]. An important question on the use of kinematic features derived from acoustic correlates (F_1 and F_2) is to which extent formant dynamics can be associated to articulation kinematics (positions and velocities of the jaw-tongue center of masses).

The assessment of AKV as a reliable kinematic correlate of articulation is carried on the multiple signal recording framework described in Fig. 2.

Fig. 2 : Speech, sEMG and Accelerometry recording.

A diadochokinetic exercise is carried on, consisting in the fast and continuous repetition of the diphthong [aj:], at a rate of 2-3 repetitions per second. Inverse adaptive filtering is used to estimate the vocal tract transfer function from running speech in real time [3]. F_1 and F_2 are evaluated from the transfer function. Surface electromyography on the masseter (sEMG) and three-channel accelerometry (Acc) are recorded also synchronously with speech. Sampling rates of sEMG and Acc are equalized to 500 Hz, as well as formant estimates, which are produced each 2 ms. The validation of formant dynamics to represent kinematic variables is done by regression of signals on the following relational chain: $sEMG > f_m$, $yAcc > \Delta y$ and $\Delta y > \Delta F_1$, ΔF_2. The results of regression studies are given in Table 1.

Table 1 : Regression results for the diadochokinetic validation exercise. r: correlation coefficient; p: p-value; S: Spearman; P: Pearson.

Correlation	r (S)	p (S)	r (P)	p (P)
Δy_r vs f_m	0.83	<0.001	0.81	<0.001
ΔF_1 vs Δy_r	-0.89	<0.001	-0.89	<0.001
ΔF_2 vs Δy_r	0.78	<0.001	0.79	<0.001

The correlation between the masseter force estimate from sEMG (f_m) and the vertical displacement of the reference point (Δy_r) is high and statistically relevant (0.83/0.81), showing a strong relation between neuromotor activity and movement. The correlation between vertical displacement (Δy_r) and formant changes are also high and relevant, stronger and counter-related with respect to ΔF_1 (-0.89), than with respect to ΔF_2 (0.78/0.79). These results are aligned with the relationship between the phonation opening (Δy_r) and the variation of the first formant (ΔF_1).

Relating KLD to clinical scores

Estimates of the AKV probability density function have been used to evaluate the Kullback-Leibler divergence (KLD) between two different distributions [4]. A correlation study between KLD's and specific neurological scores of PD patients has been conducted. Vowel utterances [a:, i:, u:] from 60 male and 40 female PD patients within an age range of 66.3±8.6 and 69±7.7 years (respectively) have been processed and statistically modelled. Similar vowel utterances from another set of 26 male and 27 female normative subjects within an age of 65.6±8.9 and 61.8±9.1 years old (respectively) have also been processed and statistically modeled to produce a normative database. Recordings were taken at 16 kHz and 16 bits. The database was collected at St. Anne's University Hospital in Brno (Czech Republic), including also meta-information from each patient as gender, age, time since first diagnosis, scores of UPDRS-III and IV, freezing of gait (FOG), non-motor symptoms (NMSS), sleep disorders (RBDSQ), mini-mental state (MMSE), Addenbrooke's cognitive evaluation (ACE-R), Beck depression inventory (BDI), faciokinesis and phonorespiratory competence.

RESULTS AND DISCUSSION

Table 2 gives the results of comparing the KLD between each PD speaker in the male database to the scores assigned by clinical neuromotor therapists by correlation.

Table 2 : Correlations between DKL and clinical scores for the male set. r: correlation coefficient; p: p-value; S: Spearman; P: Pearson. Male set.

Score	r (S)	p (S)	r (P)	p (P)
PD duration	**0.40**	**0.0038**	**0.31**	**0.0241**
UPDRS III	-0.12	0.4397	-0.19	0.2224
FOG-Q	**0.41**	**0.0024**	**0.40**	**0.0029**
NMSS	**0.44**	**0.0013**	**0.41**	**0.0024**
RBDSQ	**0.51**	**0.0002**	**0.52**	**0.0001**
MMSE	-0.15	0.3325	-0.18	0.2620
ACE-R	0.21	0.1769	0.24	0.1352
BDI	-0.10	0.5351	-0.12	0.4314

It may be seen that KLD presents a moderate but statistically relevant correlation with respect to the time interval since first diagnosis. Correlation is also relevant with respect to freezing of gait, non-motor symptoms and sleep disorders. No relevant correlation has been found to UPDRS, mini-mental state or ACR and BDI scores.

CONCLUSIONS

Correlation studies between sEMG, Acc and speech formant dynamics reveal important cause-effect relations among these traits, which allow the definition of articulation kinematic modelling. KLD from probability densities of articulation kinematics (AKV) can be used to grade the timely evolution of neurodegenerative processes, after first diagnosis. No significant correlation has been found relative to UPDRS, as this scale does not describe speech alterations properly.

REFERENCES

[1] Jürgens U. Neural pathways underlying vocal control. *Neurosci. and Behav. Rev.* 2002; 26: 235-258.

[2] Gómez A, de Arcas G, Gómez P, Álvarez A, López JM. Estimating Facial Neuromotor Activity from sEMG and Accelerometry for Speech Articulation. *Proc. of the IEEE Int. Symp. on Medical Measurements and Applications* 2018: 287-292.

[3] Deller JR, Proakis JG, Hansen JHL, *Discrete-Time Processing of Speech Signals*, NewYork, Macmillan 1993.

[4] Georgiou T, Lindquist A, Kullback-Leibler Approximation of Spectral Density Functions. *IEEE Trans. on Information Theory* 2003; 49(11): 2910-2917.

ACCURACY OF THE COMMERCIALLY AVAILABLE EVENTIDE ECLIPSE TO PERTURB AUDITORY FEEDBACK OF FUNDAMENTAL FREQUENCY

Elizabeth S. Heller Murray[1], Ashling A. Lupiani[1], Katharine R. Kolin[1,2], Cara E. Stepp[1,3,4]

[1] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA
[2] Undergraduate Program for Neuroscience, Boston University, Boston. MA
[3] Department of Biomedical Engineering, Boston University, Boston, MA
[4]Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA
Keywords: Voice; Vocal Motor Control; Acoustic; Auditory Feedback

OBJECTIVES

Measuring the vocal response to surreptitiously shifted fundamental frequency (f_o) feedback provides insight into vocal motor control. However, unintended changes to the speech signal (e.g., formants, intensity) may co-occur with the experimental shift of f_o, that is, the *pitch shift*. These changes can significantly impact the interpretability of the data or influence the participant's vocal response. This study details the effects of pitch shifting with the commercially available "Eventide Eclipse" hardware.

METHODS

Input signals of sustained vowels were produced by 10 vocally healthy participants (M = 22 years, 6 females, 4 males). Participants produced sustained vowels for approximately 3 seconds. Output signals, created in near real-time by the Eventide Eclipse, were either identical to the input, shifted down by 100 cents, or shifted up 100 cents. Pitch shifts were either initiated prior to voice onset and held constant throughout the entire utterance (*long* duration) or initiated at a variable point after voice onset and held for a 200 ms period (*short* duration). Both the intended manipulations set by the user (magnitude and duration of pitch shift) as well as the potential unintended consequences of the paradigm (F1 and F2 formant frequency changes, intensity changes, and delay between input and output) were analyzed.

RESULTS

Intended, pitch shift accuracy

The average magnitude of both short and long duration pitch shifts were within 1 cent of their intended shift. A two-way repeated measures analysis of variance (ANOVA) found significant main effects of shift duration (long, short) and magnitude (+100, -100). Trials that were shifted +100 cents were significantly less accurate than trials shifted -100 cents, with an average of 0.5 and 0.4 cents away from the intended pitch shift, respectively. Trials in which the pitch was shifted for a short duration were significantly less accurate than trials shifted for a long duration, 0.7 and 0.2 cents away from intended pitch, respectively.

Intended, pitch shift duration (short duration trials)

No significant differences ($p>0.05$) were found between short duration shifts of either +100 and -100 cents; therefore, these pitch shift magnitudes were collapsed for further analysis. The measured average duration was 11.8 ms less than the intended duration of 200 ms, that is, a 5.9% reduction in duration. This measurement included onset and offset times before and after the pitch shift reached steady state. Further analysis solely on the steady state portion of the pitch shift revealed an average of 43.3 ms (21.7%) reduction in the duration of the pitch shift.

Unintended, delay between input and output

The overall average delay of the system across trials was 11.1 ms. Trials with a +100 cent pitch shift (M = 19.4 ms, SD = 9.1 ms) had a significantly longer delay than both trials with a -100 cent pitch shift (M = 10.4 ms, SD = 5.7 ms) and trials that were not shifted (M = 10.3, SD = 7.0 ms). There were no significant differences between trials with a -100 cent pitch shift and trials that were not shifted.

Unintended, change in intensity

Measurement of intensity differences between input and output signals revealed there were no significant differences between long or short duration trials. Two paired sample t-tests found no significant differences between changes in intensity in trials that were not pitch shifted compared to trials that were pitch shifted (all $p>0.025$).

Unintended, percent change in formants (F1, F2)

Changes in F1 and F2 followed the direction of the pitch shift, either upward or downward. Examination of percent change of F1 and F2 indicated there were no significant differences between long and short duration trials. The average percent changes of F1 and F2 were 6.5% and 6.0%, respectively. This percent change of F1 and F2 could change the perception of the produced vowel in some cases.

CONCLUSIONS

The Eventide Eclipse is an accurate pitch shifter, with the average magnitude of both short and long duration pitch shifts measuring within 1 cent of their intended shift. However, when short (200 ms) duration shifts are intended, the actual implemented duration is shorter by 11.8 ms. Average delay between input and output signals was under 20 ms for all trial types. No intensity changes were noted during pitch shifts of either +100 or -100 cents. Due to the method of pitch shifting, changes F1 and F2 occurred in that direction of the pitch shift, that is, they increased for +100 cent pitch shifts and decreased for -100 cent pitch shifts. The percent change in F1 and F2 were small (6.6 % and 6.0%, respectively), yet they may be large enough to change the perception of the vowel in some cases. Overall, this hardware provides an easy-to-use, accurate method of manipulating pitch, allowing examination of vocal motor control.

ACKNOWLEDGEMENTS

EFFECT OF SUBGLOTTIC STENOSIS ON ACOUSTICAL OUTPUT OF SYNTHETIC VOCAL FOLD MODELS

Benjamin Hilton, Scott L. Thomson

Brigham Young University, Provo, UT, USA

Keywords: Subglottic Stenosis; Synthetic Vocal Fold Modeling

INTRODUCTION

Subglottic stenosis (SGS) is characterized by a narrowing of the trachea at, or slightly below, the level of the cricoid cartilage ring. The extra growth is frequently comprised of scar tissue and can be caused by prolonged intubation or trauma [1], although it is often idiopathic.

SGS causes dyspnea, and surgical management is frequently necessary. SGS is also believed to adversely affect voice quality. While significant research has been conducted on the effect of SGS on respiration, relatively few studies on its effect on voice production have been performed. Smith and Thomson [2] simulated the effect of SGS on vocal fold vibration in a two-dimensional computational fluid dynamics (CFD) simulation. They found that VF vibration and airflow were only significantly affected at around 90% occlusion and greater. Brouns et al. [3] also performed a CFD simulation on the flow dynamics of a stenosis. They found that the simulated pressure drop over a stenosis was only significant at above 70% constriction. Concerning human populations, Hillel et al. [4] analyzed the change in voice quality of 27 subjects with laryngotracheal stenosis (including both subglottic and tracheal stenosis) before and after a balloon dilation operation to restore airway patency, and found that 19 of the 27 had improved Voice-Related Quality of Life (V-RQOL) scores, five had worse V-RQOL scores, and three remained unchanged. Thus, it is clear that voice quality is related to SGS and its treatment.

The aim of this research was to experimentally explore the effects of SGS on phonation through the use of synthetic vocal fold (VF) models. This was accomplished by developing a replica of the airway, including self-oscillating vocal folds and an adjustable model of SGS in the subglottal area. Phonation was simulated with varying degrees of SGS, and metrics such as fundamental frequency and acoustic spectra were used to determine the effect of the SGS.

METHODS

Self-Oscillating Vocal Fold Models

Advantages of synthetic VF models include repeatability and ease of access. The VF models used in this research are generally based on those developed by Murray et al. [5] and consist of four layers of silicone, each layer with different material properties. The geometry of the models used in this research differs slightly from that of those in [5] in order to maintain their reasonable emulation of the vibrational properties of human vocal folds in terms of frequency, motion, and onset pressure, while also working towards a model with improved closed quotient. The revised geometry (Fig. 1) was derived from a computational optimization study of VF vibration.
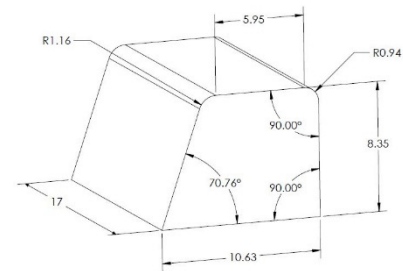
Figure 1: Outer dimensions (mm) of VF model

SGS Model

A device was developed (Fig. 2) to model SGS with varying degrees of constriction. The microcontroller-based device is actuated by stepper motors. The SGS model is constricted by pulling threads that are wrapped around a silicone sleeve. The silicone sleeve, through which the air flows to the VF models, has an inner diameter of 20 mm at its relaxed position. The sleeve can constrict until it closes off completely, meaning that for a tracheal diameter of 20 mm, the device can simulate 0% to 100% stenosis.
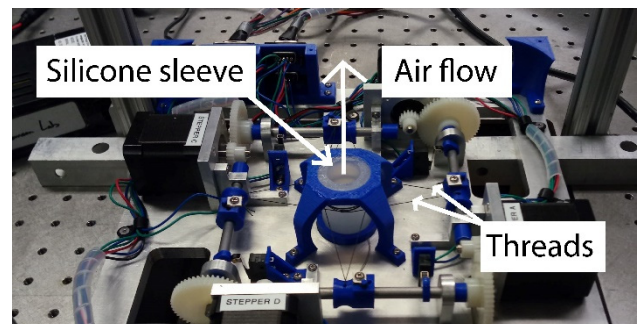
Figure 2: Experimental setup for exploring the influence of SGS on synthetic VF model response and acoustic output. VF model not shown.

Instrumentation and Measurement

The synthetic VF models were placed above the device. The stenosis was approximately 3 cm below the models. Data were collected using a PCB 377B02 free-field microphone placed 18 cm off-axis from the VFs and an NI 9234 module. Sound pressure data were recorded in LabVIEW and exported to MATLAB. Data were recorded from 0% to 100% stenosis in increments of 20%, and between 80% and 100% in increments of 5%. The order was randomized to prevent bias. The flow was adjusted for each level to maintain a subglottal pressure near 1.9 kPa, which corresponded to approximately 20% above the onset pressure of 1.6 kPa. The flow tube was constricted 30 cm upstream of the device to reduce potential acoustic effects.

RESULTS AND DISCUSSION

Unweighted sound pressure level (SPL) and fundamental frequency values were calculated for each level of constriction and are shown in Fig. 3. The SPL only fluctuated within about ±1 dB through 85% stenosis, but then decreased significantly (by about −3 dB) at 90% stenosis. The model did not vibrate at 95% and 100% stenosis. The fundamental frequency also remained relatively constant through 85% stenosis, but at 90% stenosis increased by approximately 3 Hz.

The calculated acoustic spectra for 0% and 90% constriction are shown in Fig. 4. Generally, as the constriction increased, the noise increased and the harmonics weakened.
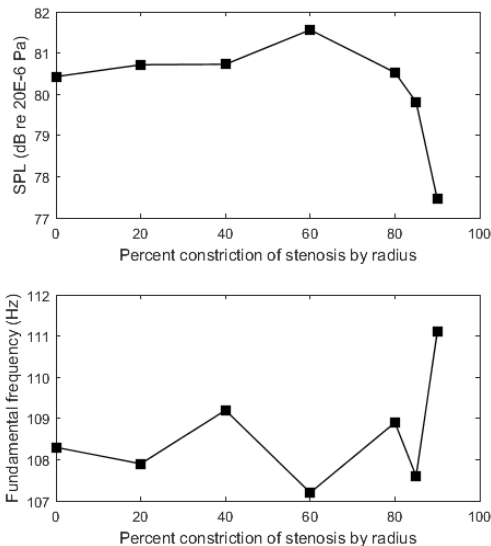
Figure 3: Sound pressure level (top) and fundamental frequency (bottom) vs. percent stenosis

CONCLUSIONS

The SPL was greatest at approximately 60% constriction, although fluctuations in SPL between 0 and 85% stenosis were relatively minor. The SPL dropped off significantly at 90% stenosis. The radiated acoustic spectral content also

varied significantly, with greater percentages of constriction causing more noise and weaker harmonics.

It is anticipated that this SGS device will facilitate study into the effects of SGS on voice production. Future work will include implementing a vocal tract to determine how SGS affects vowel sounds.
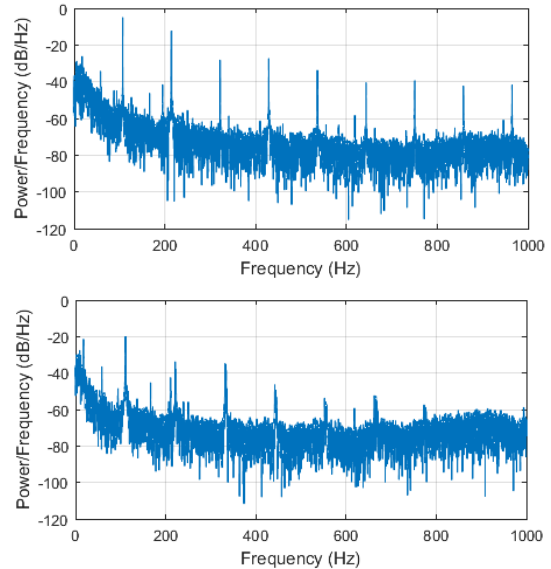
Figure 4: Acoustic spectra for 0% constriction (top) and 90% constriction (bottom).

REFERENCES

[1] Zaghi, S., Alonso, J., Orestes, M., Kadin, N., Hsu, W., Berke, G. (2016). "Idiopathic subglottic stenosis: A comparison of tracheal size," Ann Otol Rhinol Laryngol 125(8):622-626.

[2] Smith, S.L., Thomson, S.L. (2013). "Influence of subglottic stenosis on the flow-induced vibration of a computational vocal fold model," J Fluids Struct (38):77-91.

[3] Brouns, M., Jayaraju, S.T., Lacor, C., et al. (2007). "Tracheal stenosis: a flow dynamics study," J Appl Physiol 102(3):1178-84.

[4] Hillel, A.T., Karatayli-Ozgursoy, S., Benke, J.R., et al. (2015). "Voice quality in laryngotracheal stenosis," Ann Otol Rhinol Laryngol 124(5):413-418.

[5] Murray, P.R., Thomson, S.L. (2012). "Vibratory responses of synthetic, self-oscillating vocal fold models," J Acoust Soc Am 132(5):3428-38.

APPLICATION OF AUTOMATIC SPEECH RECOGNITION TECHNOLOGY FOR DESCRIBING INTELLIGIBILITY DEFICIT IN DYSPHONIC SPEECH

Keiko Ishikawa[1], Teksong Eap[1], Marissa Happ[1], Lauren Kaiser[1]

[1] Department of Speech and Hearing Science, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

Keywords: Voice; Intelligibility; Automatic Speech Recognition; Machine Learning

INTRODUCTION

Speakers with dysphonia frequently report reduced intelligibility in the presence of background noise [1]. Conventionally, clinicians measure intelligibility by conducting error analyses on manually transcribed spoken words and sentences. This process is not only highly time-consuming and laborious, but also prone to poor inter-rater reliability. The use of acoustic markers has been proposed as a solution to address these issues; however, the search of such markers has been difficult. We propose the use of automatic speech recognition technology (ASR) as an alternative approach. Recent advances in the ASR technology have resulted in the surge of applications that allows highly accurate speech-to-text transcription. We have developed an automatic speech analysis program based on IBM Watson speech-to-text application programming interface (API). This study examined the feasibility of evaluating intelligibility deficit in dysphonic speech using this approach.

METHODS

Participants

Speakers of this study were 18 speakers with normal voice and 18 speakers with dysphonic voice (6 adult females, 6 adult males, and 6 children in each group). The adults were between 21-60 years of age, and children were between 6-12 years of age. Listeners of this study were 45 native speakers of American English with normal hearing. Their hearing status was screened by pure-tone audiometry at 500, 1000, 2000, and 4000 Hz. They were considered eligible for the study when they were able to respond to the stimuli at 25 dBHL.

Stimuli Preparation

The speakers recorded 6 sentences from the Consensus of Auditory-Perceptual Analysis of Voice [2] in a single-wall sound proof booth with a unidirectional microphone (Neumann, TLM 103), placed at the microphone to mouth distance of 15 cm and 45° off axis. The recordings were digitized at a sampling rate of 44.1 kHz with a solid-state recorder (TASCAM SS-R200). The intensity of the samples was standardized at 60 dB SPL using PRAAT. After the standardization, cafeteria noise (Auditec, St Louis, Missouri, USA) was added to the samples at SNR -3, +0, and +3 with a custom-made MATLAB program.

Instrumentation and Measurement

The speech recordings were analyzed using a newly developed speech analysis program, "Speech Error." This

Original text: The <u>blue spot</u> is on the key again
Transcription: The <u>boat</u> is on the key again

$$\text{Error rate} = \frac{Number\ of\ missed\ words}{Total\ number\ of\ words} * 100 = 25\ \%$$

program utilizes IBM Watson speech-to-text service application programming interface (API) for transcribing speech and calculates error rate by comparing the transcription and the original text. The error rate was calculated as a ratio between the number of missed words and the total number of words in the original text. An example of the error calculation is shown below:

The program also evaluates the similarity between the original text and transcription using the Python SequenceMatcher, which is based on the "gestalt pattern matching" algorithm [3]. Additionally, the program retrieves the confidence level of transcribed words and number of alternative words for each transcribed word from the service and calculates their average values across words in a sentence.

The program performed quite poorly when it transcribed stimuli with the background noise, to the extent that it returned no transcription for some of the stimuli. This is a well-known problem with automatic speech recognizers [4]. For this reason, we used stimuli without the noise for the transcription.

Listening Experiment

The listeners rated intelligibility of 18 speakers with dysphonic voice on a 7-point scale in the presence of cafeteria noise at SNR-3, +0 and +3. An age/gender-matched normal sample was coupled with a dysphonic sample to provide an anchor for the listeners. The stimuli were presented via headphones (Sennheiser HD 380 pro) at 73 dB SPL.

Analysis

The difference in ASR-based measures between normal and dysphonic groups was examined with two-way Analysis of Variance (ANOVA). The association between the ASR-based measurements and averaged intelligibility

ratings across listeners was examined with the Spearman-rank correlation test.

RESULTS

Two-way ANOVA indicated that speaker's age/gender and diagnosis of dysphonia significantly affects all measures. The error rate and the average number of alternative words were significantly lower in the dysphonic group. On the other hand, the similarity and the average confidence level across words were significantly higher in the dysphonic group. The test statistics and results are shown in table 1 and figure 1.

Table 1: Results of two-way ANOVA.

Age and Gender	
Error rate	$F(2, 54) = 8.16$, $p < 0.001$
Similarity	$F(2, 54) = 8.84$, $p < 0.001$
Confidence	$F(2, 54) = 10.34$, $p < 0.001$
Alternative words	$F(2, 54) = 9.28$, $p < 0.001$
Diagnosis	
Error rate	$F(1, 54) = 4.28$, $p = 0.03$
Similarity	$F(1, 54) = 5.15$, $p = 0.03$
Confidence	$F(1, 54) = 9.59$, $p < 0.001$
Alternative words	$F(1, 54) = 5.422$, $p = 0.02$

Fig. 1: Bar plots showing the difference in error rate, similarity, average number of alternative words, and average level of confidence level between normal and dysphonic speech. The error bars indicate the standard error.
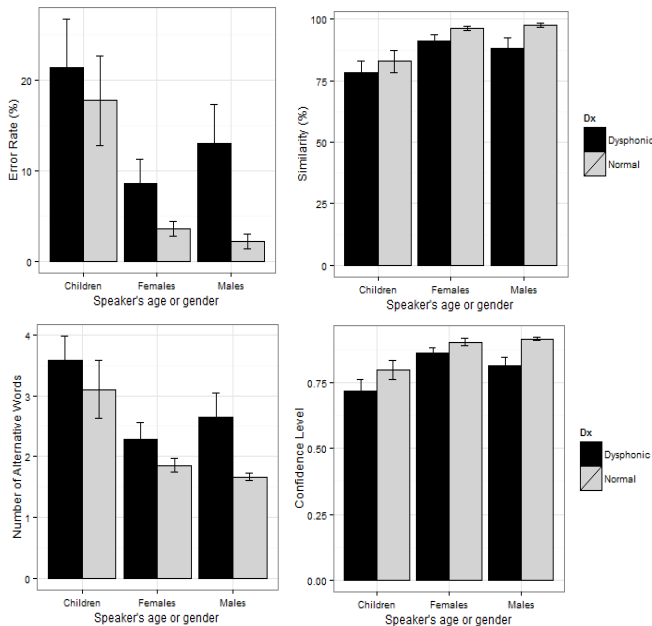
The Spearman Rank correlation tests indicated that he ASR-based measures were significant predictors for the intelligibility ratings, except for the error rate and similarity scores at SNR-3. The correlation was the strongest at SNR+3 (Table 2).

Table 2: Rho values from the Spearman Rank correlation tests. Asterisks indicate the significance level: * is p < 0.05, and ** is p < 0.01.

Noise Level	SNR-3	SNR+0	SNR+3
Error Rate	-0.44	-0.64**	-0.71**
Similarity	0.45	0.66**	0.75**
Alternative words	-0.57*	-0.71**	-0.85**
Confidence	0.52*	0.67**	0.77**

DISCUSSION & CONCLUSION

Results of the study have shown that the automatic speech recognition technology is a potential approach for describing the difference between normal and dysphonic speech, and intelligibility deficit in dysphonic speech. The difference in ASR-based measurements between normal groups suggests the system is best at transcribing adult male speech. This observation indicates that the effect of gender and age needs to be considered in future studies.

Despite the positive findings, it should be noted that the use of "off-the-shelf" algorithm is limiting as it does not inform us what influenced the measurements. Defining acoustic correlates for disordered speech is of a clinical and scientific interest. Creating our own speech recognition system is required for this type of inquiry.

Limitations of the study include that the use of rating scale for measuring intelligibility. Because rating is based on an impression of intelligibility, it is not the exact measure of the amount of information received by listeners. The future studies need to examine the correlation between transcription-based intelligibility measurement and ASR-based measurement.

REFERENCES

[1] Ishikawa, K., Boyce, S., Kelchner, L., Powell, M. G., Schieve, H., de Alarcon, A., & Khosla, S. (2017). The effect of background noise on intelligibility of dysphonic speech. *JSLHR* 2017; 60(7): 1919-1929.

[2] Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *AJSLP* 2009; 18(2):124-132.

[3] "7.4. Difflib — Helpers For Computing Deltas — Python 2.7.13 Documentation". Docs.python.org. N.p., 2017. Web. 29, November. 2017

[4] Li, Jinyu, et al. An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.4 (2014): 745-777.

FLOW-STRUCTURE INTERACTION SIMULATION OF A CANINE LARYNGEAL MODEL

Weili Jiang[1], Xudong Zheng[1], Qian Xue[1], Xiaojian Wang[1], Liran Oren[2], Charles Farbos De Luzan[2], Ephraim Gutmark[3], Sid Khosla[2]

[1] Mechanical Engineering Department, University of Maine, Orono, Maine, USA
[2] Department of Otolaryngology–Head and Neck Surgery, University of Cincinnati, Cincinnati, Ohio, USA
[3] Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Cincinnati, Ohio, USA

Keywords: Modelling; Flow-Structure Interaction; Canine Larynx

INTRODUCTION

Previous excised canine laryngeal experiments [1–3] revealed the skewing of flow rate and the presence negative pressures inside a glottis without a supraglottal vocal tract. The hypothesis was that the local negative pressure was caused by intraglottal vortices generated from flow separation in the divergent glottal channel.

The objective of this study was to: (1) develop a computational flow-structure interaction model to reproduce the dynamics of glottal airflow and vocal fold vibrations in a subject-specific realistic canine laryngeal configuration; (2) analyze the influence of intraglottal vortices and their resulting negative pressures on flow rate and vocal fold vibrations.

METHODS

A sharp-interface immersed boundary method based incompressible flow solver [4] and a finite element based solid mechanics solver were utilized to model the dynamics of airflow and vocal fold vibration, respectively. These two solvers were explicitly coupled through a Lagrangian interface where vocal tract and vocal folds contacted [5]. The geometric structure of the vocal folds and vocal tract was reconstructed from the MRI scan of a mongrel canine. The vocal folds were assumed to be a two-layer structure with transversely isotropic material properties and an inferior-superior stiffness gradient.

The material parameters were reversely determined by matching the stress-strain curve in numerical indentation to that in experimental indentation [6]. The numerical indentation [7] followed the experimental indentation by using the exactly same indentation locations and loading values. [6]. A genetic algorithm based optimization was then applied to minimize the difference of the deformations from the numerical indentation and experiment indentation. The obtained transverse Young's modulus of the body layer was 4.8kPa. The transverse Young's modulus of the cover layer decreased linearly from 3.0 kPa to 1.3 kPa from the inferior to superior due to the stiffness gradient.

The simulation setup also followed the excised experiment. Fig.1 shows the geometry of the vocal folds as well as the simulation setup. The vocal folds were mounted inside a large enclosure without a supraglottal vocal tract. A 2.0kPa pressure drop was applied between the inlet of the vocal tract and the outlet of the fluid domain. The vocal fold surface that contacted with the vocal tract was free to move and the other part was constrained not to move.
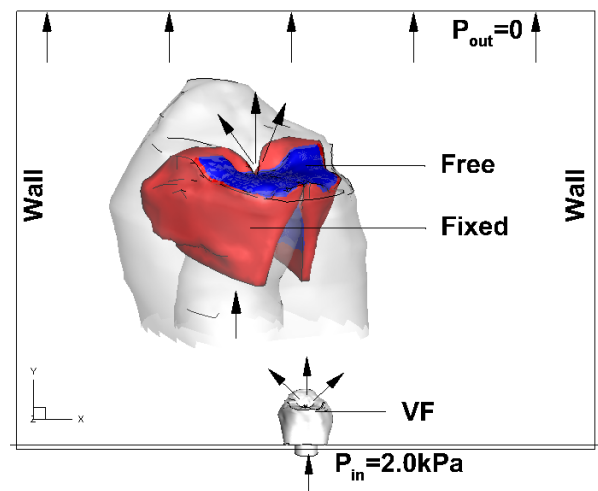
Fig. 1: Simulation setup.

RESULTS AND DISCUSSION

Table 1: Simulation results and the comparison with the experiment [3].

	F_0 (Hz)	P_{neg} (kPa)	DA (°)
Simulation	187	-1.03	41
Experiment	84 ~ 193	-0.48 ~ -0.98	49

A range of pressure drop of 2.0±0.2kPa was applied in the experiments [3]. Several key parameters including the fundamental frequency (F_0), the most negative intraglottal pressure (P_{neg}) and the divergent angle (DA) at the moment of P_{neg} were listed in Table 1 for comparison. For DA, only one value was available from the experiments and it corresponded to the pressure drop of 2.1kPa [3]. The simulation results were averaged over cycles. It was

noticed that all the parameters from the simulation were in the range or very close to the experimental results. The F_0 in the simulation was at the high end of the experiment results while the P_{neg} was at the low end.
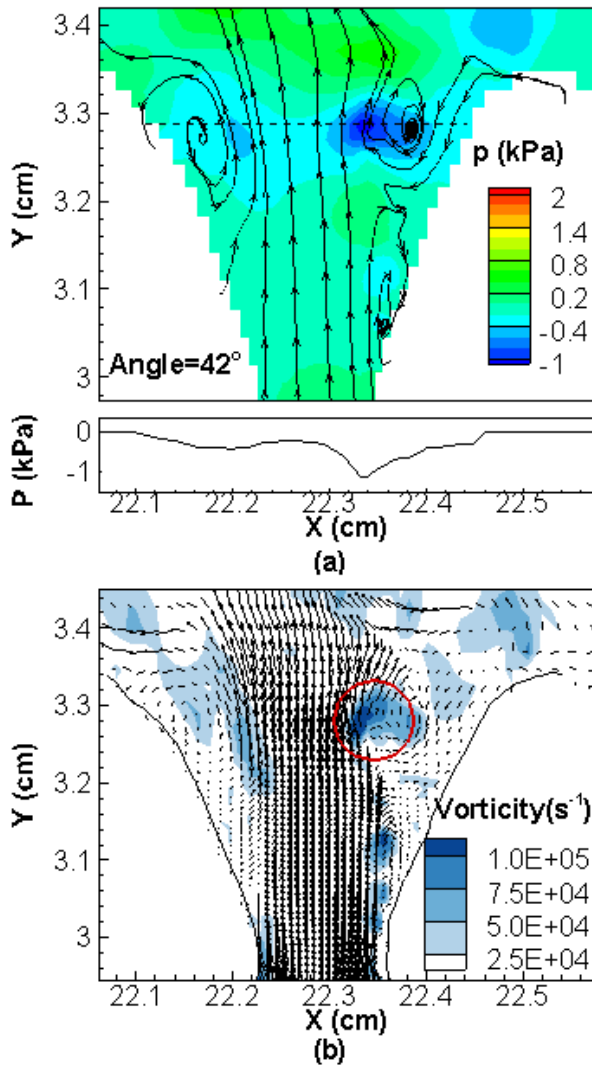
Fig.2: Intraglottal field at the instant of the most negative pressure occurred: (a) Pressure contour. The pressure value along the dashed line is shown below. (b) Vorticity with velocity vector superimposed.

Fig.2 shows the flow field at the instant when the most negative pressure occurred. Fig.2 (a) shows the pressure contour with the streamline superimposed. The glottal angle at this instant was 42°. The pressure profile along the dash line is shown below. Fig.2 (b) shows the vorticity contour with the velocity vector superimposed. It was observed that the flow separated near the inferior edge of the vocal folds. A series of vortices formed between the glottal jet and the vocal fold walls. The formation of the vortices was due to the entrainment of the glottal jet to the surrounding flow and the downstream flow entering the

space between the glottal jet and the vocal fold walls. The vorticity has the largest value near the superior edge which was denoted by a red circle in Fig.2 (b). It corresponded to the location of the lowest pressure in Fig.2 (a). The pressure profile along the dash line in Fig.2 (a) shows that the pressure decreased to -1.15 kPa at the location of the vortex center while at other locations the pressure was about -0.3 kPa. It can be seen that the low pressures of the vortices also caused negative pressures on the vocal fold walls.

CONCLUSION

A computational model for reversely determining the material properties and simulating the flow-structure interactions of vocal fold vibrations in subject-specific realistic laryngeal configurations was developed. The method was demonstrated and validated on a canine laryngeal model. Several key parameters were compared with the excised experiments and a good agreement was achieved. The results of the flow field revealed that the flow separation in the divergent glottal shape resulted in the formation of intraglottal vortices. These vortices generated significantly lower negative pressures at the vortex centers which also caused negative pressures on the vocal fold walls.

ACKNOWLEDGMENTS

REFERENCES

[1]   L. Oren, S. Khosla, E. Gutmark, J. Acoust. Soc. Am. 135 (2014) 380–388.

[2]   L. Oren, E. Gutmark, S. Khosla, J Acoust Soc Am 137 (2015) 935–943.

[3]   L. Oren, S. Khosla, E. Gutmark, J. Biomech. 47 (2014) 1287–1293.

[4]   R. Mittal, H. Dong, M. Bozkurttas, F.M. Najjar, A. Vargas, A. von Loebbecke, J. Comput. Phys. 227 (2008) 4825–4852.

[5]   X. Zheng, Q. Xue, R. Mittal, S. Beilamowicz, J. Biomech. Eng. 132 (2010) 111003.

[6]   L. Oren, D. Dembinski, E. Gutmark, S. Khosla, J. Voice 28 (2014) 297–304.

[7]   B. Geng, Q. Xue, X. Zheng, J. Acoust. Soc. Am. 141 (2017) EL351-EL356.

THE EFFECT OF PROTEIN COATING ON EPITHELIAL BARRIER INTEGRITY, CELL PROLIFERATION, AND CELL PHYSIOLOGY IN PRIMARY CULTURE OF VOCAL FOLD EPITHELIAL CELLS

Emily E. Kimball MS[1], Carol Xu MD PhD[2], Leanne Sayce DPhil[2], Bernard Rousseau MMHC PhD CCC-SLP[1,2]

[1] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, USA.
[2] Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA.

Keywords: Vocal Fold; Epithelium; Cell Culture; Basement Membrane

INTRODUCTION

To effectively study the injury response of the vocal fold, it is important to characterize the normal structure the vocal fold epithelial barrier. The epithelium provides a structural barrier to the underlying lamina propria, and is often considered the first line of defense against the shearing forces of vibration [1]. As such, understanding the unique role of the epithelium in the wound healing process will guide future *in vivo* studies of vocal fold injury and repair. The Laryngeal Biology Laboratory at Vanderbilt University currently has a protocol for primary vocal fold epithelial cell culture using a rabbit model [2].

In the cell culture process, the choice of protein substrate is critical, because it provides the growing epithelial cells a surface with which to interact and adhere [3]. Although collagen I is present in large quantities in the vocal fold lamina propria, it is not typically present in the basement membrane [4]. Collagen IV, laminin, and fibronectin are primary components of the basement membrane and have the potential to contribute directly to cell signaling in the vocal fold epithelial cell layer. Matrigel was also tested in this study, as its primary components include laminin and collagen IV, and is a widely used basement membrane substitute *in vitro* [5]. We hypothesized that one or multiple proteins tested would facilitate more consistent proliferation and TEER in our cell culture model compared to collagen I.

METHODS

Cell Culture

Larynges (2 rabbit and 2 human) were enzymatically treated to isolate vocal fold epithelial cells. At passage 2, cells were seeded onto inserts coated with Collagen I, Collagen IV, Laminin, Fibronectin, or Matrigel. Transepithelial electrical resistance (TEER) was measured and phase contrast images were collected. Rabbit cells were fixed for immunofluorescence labeling of proliferation (Ki67) and differentiation (Vimentin).

Data Collection

P2 epithelial cells were cultured on a coated porous cell culture insert suspended in a 6-well plate. TEER was measured using an epithelial voltohmmeter. For the experiments described, TEER was measured in each well once daily from days 4-18 of P2. Phase contrast images were collected in rabbit cell line 1 on days 4 and 16 of passage 2 using a 10x objective.

Cell culture inserts from rabbit epithelial cell line 1 were fixed with methanol on days 4, 7, 10, 13, 16, 19 and 30. Cell proliferation was assessed via fluorescence immunolabeling with a monoclonal antibody against Ki67 (Abcam ab15580). Images were captured via widefield fluorescence microscopy at 20x.

Cell differentiation was assessed using immunofluorescence of Vimentin (Abcam ab45939). Images were captured using a Nikon spinning disk confocal miscroscope using 5µm optical sectioning with a 20x objective. A maximum intensity projection of each z-stack was used for quantitative analysis.

Analysis

For each protein substrate at each time point, the percent of cells exhibiting positive Ki67 staining was determined using ImageJ [6]. Percent positive Vimentin staining area was calculated for each image. Images were randomized, and an intensity threshold was manually selected using ImageJ [6]. Percent positive area was calculated based on this manual threshold. For intra-rater reliability, 10% of samples were redundantly analyzed (R^2=0.81).

Two-way analysis of variance (ANOVA) was computed across protein conditions at each time point. If there was a significant main effect of protein at any time point for any measure, a post-hoc Tukey test was performed for pairwise comparison. Statistical significance was determined using an alpha of 0.05.

RESULTS

TEER

Despite inherent variability in the rabbit cell TEER data, statistically significant trends emerged. Firstly, TEER increased over time in both rabbit cell lines. Protein substrate was also determined to significantly impact TEER. In both rabbit epithelial cell lines, Laminin elicited the lowest TEER values, and conversely, Collagen IV consistently elicited the highest TEER values. Collagen I,

a commonly used protein substrate for cell culture, yielded moderate TEER values (Fig. 1a,b).

Human epithelial cells were more resilient to changes in environment and appear to demonstrate greater adaptability than their rabbit counterparts. TEER across protein conditions was less variable in the human cell lines than in the rabbit cell lines (Fig. 1c,d). Because TEER was more variable and condition-dependent in the rabbit cell lines, further investigation was carried out in one rabbit epithelial cell line (Rabbit Epithelial Cell Line 1).
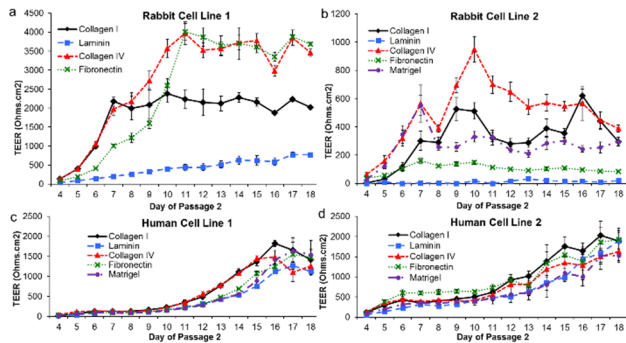

Fig. 1

Cell Morphology

At day 4, cells grown on all substrates, except Laminin, were confluent. Cells in the Collagen I, Collagen IV, and Fibronectin conditions shared round, cobblestone cell morphology. Cells in the Laminin condition presented with vastly different morphology, with long, spindle-like cell shapes.

Cell Proliferation

Cells grown in the Collagen IV condition trend toward having a higher percent of cells positive for Ki67 at later time points, but a greater number of samples is needed to discern if this is a genuine difference (Fig. 2a).

Cell Differentiation

At days 4, 13, 16 and 19, the Laminin condition had significantly elevated Vimentin labeling relative to other substrates ($p<0.05$; Fig. 2b).
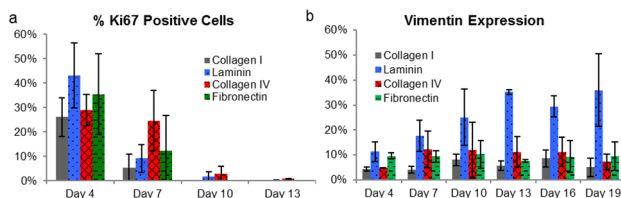

Fig. 2

DISCUSSION

The interaction of vocal fold epithelial cells with the protein components of the basement membrane *in vivo* is complex. There are many epithelial membrane-bound proteins (e.g. integrins) that have the potential to capture chemical and tactile signals from the extracellular protein scaffold and subsequently translate those signals into changes within the epithelial cell barrier. This provides a robust cellular mechanism that can regulate a wide range of functions and behaviors depending on the local environment, including cell adhesion, cell morphology, cell differentiation, and epithelial barrier integrity [7]. While not all of these mechanisms were expressly demonstrated in the current study, the data presented support a major role of basement membrane proteins in the integrity of the isolated epithelial cell layer *in vitro*.

CONCLUSION

Continued investigation is required to further elucidate the physiology of vocal fold epithelial cells *in vitro* and how this behavior translates to an *in vivo* environment. However, in the current study, Collagen IV appears to provide the most robust cell culture substrate for vocal fold epithelial cells across both rabbit and human species.

REFERENCES

[1] CK Novaleski, EE Kimball, M Mizuta, and B Rousseau, "Acute exposure to vibration is an apoptosis-inducing stimulus in the vocal fold epithelium," *Tissue Cell*, vol. 48, no. 5, pp. 407–416, Oct. 2016.

[2] M Mizuta, T Kurita, EE Kimball, and B Rousseau, "Structurally and functionally characterized in vitro model of rabbit vocal fold epithelium," *Tissue Cell*, vol. 49, no. 3, pp. 427–434, Jun. 2017.

[3] JC Adams and FM Watt, "Expression of beta 1, beta 3, beta 4, and beta 5 integrins by human epidermal keratinocytes and non-differentiating keratinocytes," *J. Cell Biol.*, vol. 115, no. 3, pp. 829–841, Nov. 1991.

[4] VS LeBleu, B MacDonald, and R Kalluri, "Structure and Function of Basement Membranes," *Exp. Biol. Med.*, vol. 232, no. 9, pp. 1121–1129, Oct. 2007.

[5] CS Hughes, LM Postovit, and GA Lajoie, "Matrigel: A complex protein mixture required for optimal growth of cell culture," *PROTEOMICS*, vol. 10, no. 9, pp. 1886–1890, May 2010.

[6] CA Schneider, WS Rasband, and KW Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 671–675, Jul. 2012.

[7] MS Liberio, MC Sadowski, C Soekmadji, RA Davis, and CC Nelson, "Differential Effects of Tissue Culture Coating Substrates on Prostate Cancer Cell Adherence, Morphology and Behavior," *PLoS ONE*, vol. 9, no. 11, Nov. 2014.

IMAGING AND QUANTIFYING DEHYDRATION AND REHYDRATION IN VOCAL FOLD TISSUE LAYERS

Renee E. King[1,2], Kevin Steed[3], Ana Estefania Rivera[1,4], Jonathan J. Wisco[3], Susan L. Thibeault[1,2]

[1] Division of Otolaryngology, Department of Surgery, University of Wisconsin-Madison, Madison, Wisconsin, USA
[2] Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA
[3] Department of Physiology and Developmental Biology, Brigham Young University, Provo, Utah, USA
[4] School of Medicine, Ponce Health Sciences University, Ponce, Puerto Rico, USA

Keywords: Voice; Hydration; Laryngeal imaging

INTRODUCTION

Clinicians commonly recommend increased hydration to patients with dysphonia. However, effects on clinical voice outcome measures have been inconsistent. Hydration-induced change within different layers of vocal fold tissue is currently unknown. Magnetic resonance imaging (MRI) is a promising method of noninvasively measuring water content in vocal folds. The objective of this study was to image and quantify changes in water content within vocal fold mucosa and thyroarytenoid muscle after dehydration and rehydration.

METHODS

Excised porcine larynges (n = 30) were imaged using proton density (PD) weighted MRI at baseline and after 30 minutes immersion in deionized water (H2O), phosphate-buffered saline (PBS), 5%, 10%, or 30% sodium chloride (NaCl), or dry air. Larynges dehydrated in hypertonic solutions and dry air were also imaged after 30 minutes rehydration in H2O. Normalized signal intensity was measured in mucosa and thyroarytenoid muscle.

RESULTS

Scans revealed fluid-rich vocal fold mucosa that was distinct from muscle at baseline. Baseline normalized signal intensity in mucosa and muscle varied by left vs. right vocal fold ($p<0.01$) and by anterior, middle, or posterior location ($p<0.0001$). Intensity changes in the middle third of vocal fold mucosa differed by solution after immersion ($p<0.01$). Hypertonic solutions dehydrated the middle third of mucosa by over 30% ($p<0.001$). No difference from baseline was found in anterior or posterior mucosa or in muscle after immersion. No association was found between intensity change in mucosa and muscle after immersion. After rehydration, intensity did not differ by solution in any tissue, and was not different from baseline, but post-rehydration intensity was correlated with post-immersion intensity in both mucosa and muscle ($p<0.05$).

CONCLUSION

PD-MRI can be used to visualize large mammalian vocal fold tissue layers and to quantify changes in water content within vocal fold mucosa and thyroarytenoid muscle independently. Degree of change in vocal fold water content induced by hypertonic solutions *ex vivo* persists after rehydration.

ACKNOWLEDGMENTS

ANALYSIS OF THE AERO-ACOUSTIC SOUND SOURCES OF PHONATION IN A SYNTHETIC LARYNX MODEL

Stefan Kniesburges[1], Alexander Lodermeyer[2], Matthias Tautz[2], Michael Döllinger[1] , Stefan Becker[2]

[1] Dep. of Otorhinolaryngology, Div. of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany
[2] Inst. of Process Machinery and Systems Engineering, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany

Keywords: Synthetic larynx models; Aeroacoustic sound sources; Glottis closure insufficiency

INTRODUCTION

People with glottis closure insufficiency (GCI) suffer from an aspirated voice. However, the primary sound generation in patients concerned cannot be investigated in-vivo owing to the restricted accessibility of the larynx. Thus, synthetic larynx models are applied enabling to reproduce normal phonation and GCI allowing for analyzing the corresponding sound generation process.

METHODS

Our synthetic larynx model includes silicone vocal folds. The model is able to reproduce normal phonation as well as GCI. The flow field downstream of the vocal folds was measured in a two-dimensional mid-coronal plane by high-speed particle image velocimetry. The subglottal pressure and sound pressure in the far field were synchronously measured. Based on the flow velocity field, the acoustic sources were calculated accordingly to Lighthill's acoustic analogy [1,2]. Details of the experimental larynx model and the measuring setup can be found in [3,4]

Fig. 1: Acoustic source field according to Lighthill's acoustic analogy [1,2]. TOP: vocal fold oscillation without glottis closure representing GCI. BOTTOM: vocal folds oscillations with glottis closure.

RESULTS

For the simulated normal phonation case, the glottal jet was periodically interrupted by the vocal folds. For the GCI case, the jet remained during the complete oscillation cycle. The sound pressure signals showed significant differences. In comparison to normal phonation, the sound for the GCI case unveiled a large decrease of the tonal sound and a drastically reduced number of higher harmonics of the fundamental frequency. The analysis of the acoustic source distribution showed that the intensity of the main source directly located downstream of the vocal folds largely decreased for the GCI case.

DISCUSSION & CONCLUSION

The tonal sound decrease in case of GCI is most likely attributed to the intensity decay of the main acoustic source immediately downstream of the vocal folds. In this region, the interaction between the glottal jet and the supraglottal fluid immediately after the glottis opening can be described as pressure impulse. In cases of GCI, this sudden interaction almost totally disappears owing to the lack of glottal flow interruption.

ACKNOWLEDGMENTS

REFERENCES

[1] Lighthill M.J.. *On sound generated aerodynamically I. General theory. P Roy Soc Lond A Mat* 1952; 211(1107): 564-587.

[2] Lighthill M.J.. *On Sound Generated Aerodynamically. II. Turbulence as a Source of Sound. P Roy Soc Lond A Mat* 1954; 222(1148): 1-32.

[3] Lodermeyer A., Becker S., Döllinger M., Kniesburges S.. *Phase-locked flow field analysis in a synthetic human larynx model. Exp Fluids* 2015; 56(4):77.1-77.13.

[4] Lodermeyer A., Tautz M., Becker S., Döllinger M., Birk V., Kniesburges S.. *Aeroacoustic analysis of the human phonation process based on a hybrid acoustic PIV approach. Exp Fluids* 2018; 59(1):1-13.
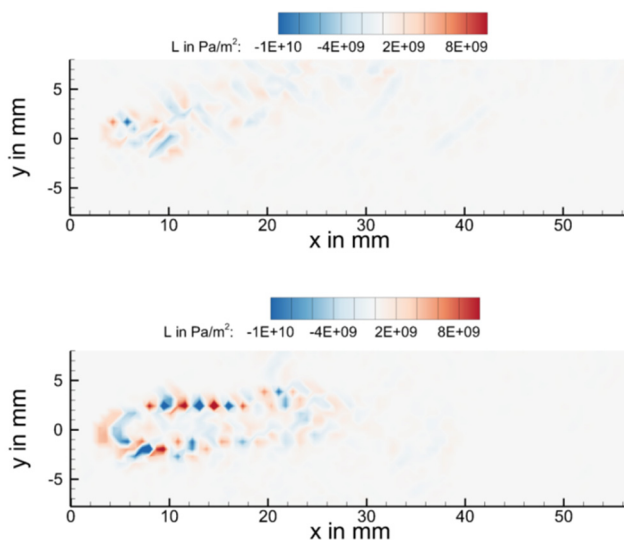
VIRTUAL HISTOLOGY WITH MULTIMODAL NONLINEAR IMAGING AND NANO-COMPUTED TOMOGRAPHY FOR QUANTITATIVE ANALYSIS OF VOCAL FOLD STRUCTURE AND INJURY

Ksenia Kolosova[1], Alexei Kazarine[2], Huijie Wang[3,4], Almoaidbellah Rammal[5], Rui Tahara[6], Karen M. Kost[5], Nicole Y. K. Li-Jessen[4,5,7], Luc Mongeau[3], Paul W. Wiseman[1,2]

[1] Department of Physics, McGill University, Montréal, Québec, Canada
[2] Department of Chemistry, McGill University, Montréal, Québec, Canada
[3] Department of Mechanical Engineering, McGill University, Montréal, Québec, Canada
[4] School of Communication Sciences and Disorders, McGill University, Montréal, Québec, Canada
[5] Department of Otolaryngology, Head and Neck Surgery, McGill University, Montréal, Québec, Canada
[6] Redpath Museum, McGill University, Montréal, Québec, Canada
[7] Department of Biomedical Engineering, McGill University, Montréal, Québec, Canada

Keywords: **Virtual Histology; Vocal Fold Scar; Nonlinear Microscopy; Microcomputed Tomography**

INTRODUCTION

Vocal fold injury and disease can lead to tissue scarring and impairment, resulting in vocal dysfunction. Scarred tissue exhibits markedly different distribution and organization of extracellular matrix components including collagen and elastin compared to healthy tissue. In histopathological analysis, these components need to be visualized to understand the scarring process and evaluate the outcomes of potential treatments such as injectable vocal fold biomaterials.

We used nonlinear laser-scanning microscopy (NLSM) supported by solvent-based tissue clearing[1], and nano-computed tomography (nano-CT) imaging, to visualize the structure of intact vocal folds and surrounding laryngeal tissue in a rabbit model. This virtual histology approach utilizes intact tissue, circumventing artifacts introduced by traditional histological slicing.

METHODS

The label-free NLSM technique uses the high second-order nonlinear susceptibility of organized collagen fibers, and two-photon autofluorescence of other tissue components including elastin fibers and blood vessels, to visualize tissue structure with high resolution in three dimensions by optical sectioning multiphoton imaging. Dissected rabbit larynges were prepared with the uDISCO protocol for tissue optical clearing[1].

Imaging was performed at 780 nm excitation using the Mira 90 Ti:Sapphire laser (Coherent) pumped by a V18 Verdi laser (Coherent).
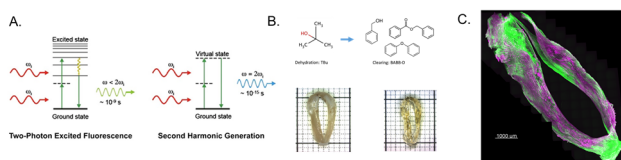
Fig. 1: NLSM methodology. Nonlinear laser-scanning microscopy directly visualizes tissue label-free, due to the (A.) two-photon excited autofluorescence (TPAF) of elastin fibers and other tissue components, and the second-harmonic generation (SHG) property of collagen fibers. Figure from [2]. (B.) To address scattering, excised tissue was optically cleared using the uDISCO method, involving sequential dehydration and subsequent immersion in a refractive index-matching solvent.[1] Resolution down to 248 nm/pixel was attained, though coarser resolution was used for larger scanning areas for tractable scanning time and data storage. (C.) 3D render of a rabbit laryngeal dissection, imaged at 795 nm/pixel in (x,y) and 50 µm in (z). Green represents SHG and magenta represents TPAF.

The CT technique uses intrinsic density differences, enhanced with phosphotungstic acid (PTA) staining[4], to image the architecture of the excised larynx down to the sub-micron level. After NLSM imaging, the optical clearing process was reversed and the dissected rabbit larynx was immersed in 70% ethanol before staining.

µ- and nano-CT were performed with the Xradia 520 Versa (Zeiss), capable of spatial resolution down to 700 nm.
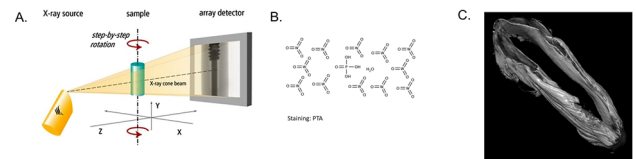
Fig. 2: CT methodology. The resolution of CT is on the micrometer scale (µ-CT) or nanometer scale (nano-CT) depending on the size of the area of interest. (A.) CT is acquired sequentially with equal resolution in (x,y,z). Figure from [3]. (B.) PTA is used as a density contrast enhancing agent.[4] (C.) 3D render of a rabbit laryngeal dissection, constructed of cubic voxels with 11 µm sides. Intensity corresponds to relative density difference.

The CT and NLSM images were visualized and rendered with the softwares Dragonfly (Object Research Systems) and Imaris (Bitplane).

RESULTS

Nano-CT and NLSM imaging visualized the structure of the vocal fold lamina propria, blood vessels, muscle, and surrounding cartilages, including visualization of chondrocytes (Fig. 3, Fig. 4).
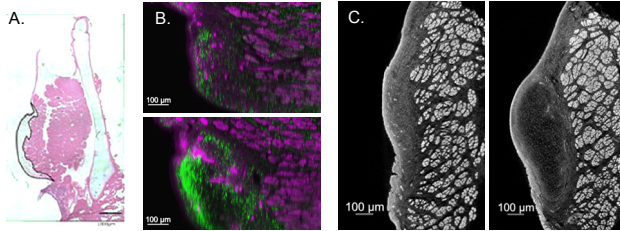
Fig. 3: Coronal slices from histology, NLSM, and nano-CT. (A.) H&E staining of a tissue slice. VF lamina propria is outlined in black. Figure adapted from [5]. (B.) Virtual slice from NLSM acquired at 640 nm (x,y) by 5 μm (z) resolution. Green represents SHG and magenta represents TPAF. Top image is sliced through vocal fold, bottom image sliced through cartilage. (C.) CT at 800 nm voxel size. Intensity corresponds to relative density difference. Left image represents vocal fold and right image represents cartilage.

NLSM also clearly resolved collagen fibers, individual blood cells, and muscle striation (Fig. 4). Larger-scale μ-CT scans (not shown) of larger laryngeal dissections provided coarse-scale information on tissue architecture, which can be used to better inform surgical procedure for injury and injection.

Fig. 4: Representative images of laryngeal tissue structures visualized with NLSM and nano-CT. In the NLSM images, green represents SHG and magenta represents TPAF, and resolution down to 248 nm is attained in (x,y). In the CT images, intensity corresponds to relative density difference, and resolution down to 800 nm is attained in (x,y,z).

CONCLUSION

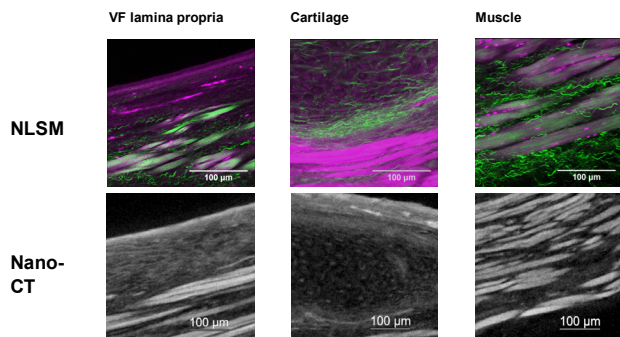NLSM and CT provide complementary information on vocal fold tissue architecture, as outlined in Table 1. Both techniques are platforms for quantitative and qualitative image analysis, and information acquired will inform treatment progression for vocal fold scarring.

Ongoing and future studies will conduct 3D segmentation for volumetric analysis of CT images, and image processing to quantify extent of scarring in NLSM images.

Table 1: Comparison of NLSM and CT parameters

	NLSM	CT
Visualization	• Direct visualization of extracellular matrix components * Collagen (SHG) * Elastin (TPAF)	• Visualization of density differences
Resolution	• Diffraction-limited in (x,y) • Loss of resolution with penetration depth	• Equipment resolution limit of 700nm • Equal in (x,y,z)
Sample preparation	• Label-free • Intact tissue requires optical clearing	• Soft tissue imaging requires contrast agent
Sample size	• Objective working distance of 2mm in (z) • (x,y) dimensions limited by stage size	• Maximum approximately 30 cm x 30 cm x 30 cm

REFERENCES

[1] Pan, Chenchen, et al. "Shrinkage-mediated imaging of entire organs and organisms using uDISCO." *Nature Methods* 13.10 (2016): 859-867.

[2] Pantazis, Periklis, et al. "Second harmonic generating (SHG) nanoprobes for in vivo imaging." *Proceedings of the National Academy of Sciences* 107.33 (2010): 14535-14540.

[3] Shenoy, Archana R., et al. "Cone-beam computed tomography." *Int Poster J Dent Oral Med*, 12.3 (2010).

[4] Metscher, Brian D. "MicroCT for comparative morphology: simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues." *BMC Physiology* 9.1 (2009): 11.

[5] Kodama, Haruka, et al. "The Ferret as a Surgical Model for Vocal Fold Scar Creation and Treatment." *Annals of Otology, Rhinology & Laryngology* (2018): 0003489417750165.

VOICE ENERGY ULITIZATION AND EFFICIENCY

Michael Krane[1], Gage Walters[1], Feimi Yu[2], Lucy Zhang[2]

[1]Applied Research Laboratory, Penn State University, State College, PA, USA
[2]Mechanical Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, USA

***Keywords*: Acoustics; Aerodynamics and kinematics; Modeling**

INTRODUCTION

This presentation describes energy flow in phonation. In order to develop a rational basis for defining voice efficiency, the energy/work inputs, outputs, and losses are defined for the phonatory airstream, using the integral mechanical energy equation (MEE). Phonatory power flows are estimated from reduced-order and high-fidelity simulations of phonation, and measurements in a physical model of the upper airway. Acoustic efficiencies are estimated from physical model experiments.

METHODS

Reduced-order model

A reduced-order model of phonation, consisting of a constant circular cross section (area $S = 5\text{cm}^2$) trachea and vocal tract, and a Story-Titze [1] larynx model, was used to estimate MEE terms. At the trachea entrance, the incoming wave pressure is $p_L = \rho c Q_L / S$, while the mouth boundary is modeled using a time-domain filter derived from a baffled circular cylinder impedance [2]. For the case shown here, p_L =700Pa.

High-fidelity simulation

The high-fidelity simulation describes a 2-D vocal fold vibration in the larynx. Fig. 1 shows the computation domain, which includes a trachea, larynx, and a vocal tract resonator, and space outside the resonator open end. To control the reflections of acoustic waves, perfectly matched layers (PML) [3] are located at the trachea entrance and on the domain boundaries outside the mouth to absorb outgoing waves. An isentropic, slightly-compressible model for air motion permitted simultaneous computation of airflow and sound. The fully-coupled aeroelastic-aeroacoustic simulation uses the modified Immersed Finite Element Method (IFEM) formulation [4,5]. A lung pressure of 1 kPa was applied in the subglottal region using a force distribution. The vocal fold was modeled in 2 layers, described by a linear visco-elastic model. Cover and body elastic moduli were set to 20kPa and 40kPa, respectively.

Physical model experiments

Physical model measurements were also performed, for 3 vocal fold models that exhibited different vibration patterns. Details are given in [5].
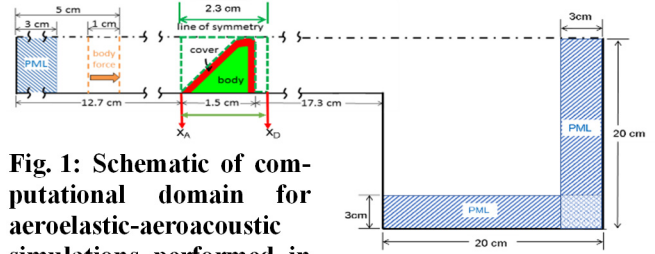
Fig. 1: Schematic of computational domain for aeroelastic-aeroacoustic simulations performed in this study and in [5]

Analysis: MEEs for a control volume

The MEE for a control volume was used as an organizing principle for analysis of the results. For the laryngeal control volume, this equation can be written as:

$$2\left(p_A^+ Q_A - p_D^- Q_D\right) = \frac{\rho c}{S} Q_D^2 + \dot{W}_{VF} + \frac{\rho c}{S} Q_A^2 + \dot{W}_{diss} + \frac{d}{dt} KE_{larynx} \quad (1)$$

$$\underbrace{}_{\text{INPUT}} \quad \underbrace{}_{\text{output}} \underbrace{}_{\text{output}} \underbrace{}_{\text{loss}} \underbrace{}_{\text{loss}} \underbrace{}_{\text{storage}}$$

The first term (work done by the driving pressure force) is identified as the work input the laryngeal control volume, while acoustic work done by laryngeal airflow on air in the vocal tract and on the vocal folds are identified as outputs. Some energy is also stored in glottal jet motion. The remaining terms denote loss mechanisms.

For a control volume containing both the larynx and the vocal tract, the MEE is given by:

$$2p_A^+ Q_A = -p_m Q_m - \dot{W}_{VF} - \frac{\rho c}{S} Q_A^2 + \dot{W}_{diss} + \frac{d}{dt} KE_{larynx} + \frac{dE_{VT}}{dt} \quad (2)$$

$$\underbrace{}_{\text{INPUT}} \underbrace{}_{\text{output}} \underbrace{}_{\text{output}} \underbrace{}_{\text{loss}} \underbrace{}_{\text{loss}} \underbrace{}_{\text{storage}} \underbrace{}_{\text{storage}}$$

where the new terms describe energy storage in vocal tract resonances, and work done by the resonant acoustic field on air outside the mouth.

Analysis: Efficiency measures

Efficiency is defined as the ratio of output to input energy. Using the inputs and outputs identified in Eqns. 1 and 2, efficiencies can be defined for the larynx (η_L), for the vocal tract (η_{VT}), and for the combined volumes (η_C), where overbars indicate average over a cycle:

$$\eta_L = \frac{\frac{\rho c}{S}\overline{Q^2}}{\overline{\Delta p Q}}, \quad \eta_{VT} = \frac{\frac{S}{\rho c}\overline{p_m^2}}{\frac{\rho c}{S}\overline{Q^2}}, \quad \eta_C = \frac{\frac{S}{\rho c}\overline{p_m^2}}{2p_L Q} \approx \eta_L \eta_{VT} \quad (3)$$

Fig. 2: Laryngeal control volume MEE waveforms, from *(top)* reduced-order model, *(bottom)* high-fidelity simulation.

RESULTS AND DISCUSSION

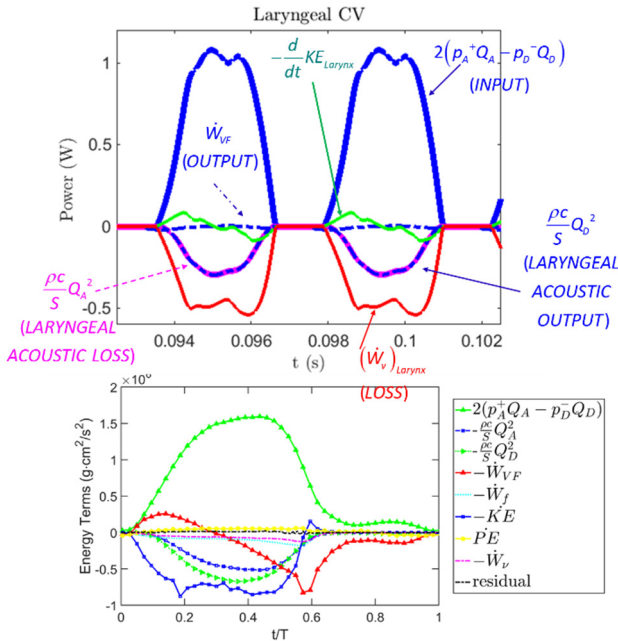Fig. 2 show waveforms of terms of the MEE, computed from output of the reduced-order and high-fidelity models, for the laryngeal control volume. Differences reflect simplifying assumptions used in the reduced-order model. These result in different vibration patterns. In addition, the glottal jet loss in the reduced-order model is larger because the jet is assumed to dissipate immediately, which doesn't occur in the high-fidelity simulation. However, the relative sizes of driving pressure input work and glottal outlet acoustic work are similar.

**Table 1: Partition of energy averaged over one cycle (reduced order model).**

LARYNX	% Input	TOTAL %input	LARYNX + VT	% Input	TOTAL %input
OUTPUTS		24	OUTPUTS		2.2
Acoustic	22.2		Acoustic	0.4	
VFs	1.8		VFs	1.8	
LOSSES		75.7	LOSSES		75.7
Jet	53.5		Jet	53.5	
Lungs	22.2		Lungs	22.2	
STORAGE		0.3	STORAGE		22.1
- larynx	0.3		- larynx	0.3	
			- VT	21.8	

Table 1 shows energy or work done per cycle for each MEE term for both the larynx and the combined volumes. It is clear that energy output from the mouth is small due to the radiation load at the mouth. The acoustic energy transfer from the larynx to the vocal tract is partitioned into storage in the vocal tract resonances and radiation from the mouth.

Fig. 3 shows efficiency measures estimated from physical model experiments [5], as a function of lung

pressure p_L. The vocal fold model exhibiting the greatest degree of closure (*Ellipse*) shows the highest efficiency for all three measures, indicating that larynx acoustic output possesses the highest energy relative to its input, and also excites the resonator most effectively. As the degree of closure decreases (from *Ellipse>M5L>M5*), efficiencies decrease as expected.

Fig. 3: Voice efficiency measures from physical model measurements [5]. (top) laryngeal efficiency η_L, (mid) vocal tract efficiency η_{VT}, (bot) combined larynx and vocal tract efficiency η_C. all efficiency measures plotted vs. lung pressure p_L.
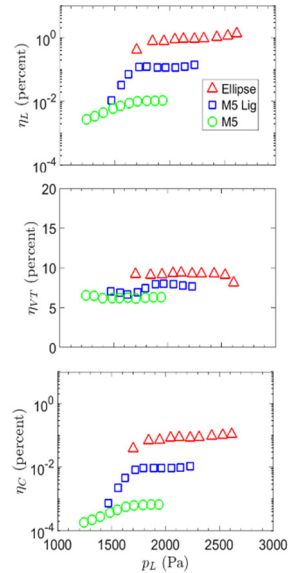
CONCLUSION

This work has introduced a physics-based theoretical framework for defining efficiency measures that distinguish between sound generation and sound transmission through the vocal tract have been developed. Simulations of phonation and physical model experiments have quantified both energy utilization mechanisms and efficiency measures.

REFERENCES

[1] Story, B. Titze, I., Voice simulation with a body-cover model of the vocal folds, *J Acoust Soc Am* 1995;97(2):1249-1260.

[2] Verge, M-P, "Aeroacoustics of confined jets, with application to the physics modeling of recorder-like instruments," PhD diss. TU Eindhoven, 1995.

[3] Yang, J., Yu, F., Krane, M., Zhang, L. The effectiveness of perfectly matched layer in simulations of fluid dynamics and fluid-structure interactions, *J Fluids and Structures*, 76: 135-152, 2018

[4] Yang, J., Wang, X., Krane, M., Zhang, L. Fully-coupled aeroelastic simulation with fluid compressibility - for application to vocal fold vibration, *Computer Methods in Applied Mechanics and Engineering,* 315:584-606, 2017.

[5] Yu, F., Walters, G., Krane, M., and Zhang, L. Phonation aeroacoustic source strengths, *Proceedings of the 11th ICVPB*, E. Lansing, MI, Aug. 2018.

AN INVESTIGATION OF VOCAL FATIGUE USING A DOSE-BASED VOCAL LOADING TASK

Zhengdong Lei[1], Laura Fasanella[1], Nicole Li-Jessen[2], Luc Mongeau[1]

[1] Department of Mechanical Engineering, McGill University, Montreal, Quebec, Canada
[2] School of Communication Sciences and Disorders, McGill University, Montreal, Quebec, Canada

***Keywords*: Vocal Fatigue, Distance Dose, Vocal Loading Task.**

INTRODUCTION

Prolonged or inappropriate voice use is regarded as one of critical factors in developing vocal fatigue [1]. Vocal loading tasks (VLTs) are often used to investigate the relationship between voice use and vocal fatigue in laboratorial settings [2]. In most of reported VLTs, participants were instructed to complete standardized reading tasks of prescribed durations at specific loudness levels. Perceptual, acoustic, and aerodynamic measurements were then used to evaluate changes in performance and quality of voice before and after VLTs [3-7]. Inconsistent or sometimes contradictory findings have been reported. One common drawback is that the prescribed VLT might not induce a detectable level of vocal fatigue across individuals. A more robust design of VLT is thus warranted in the investigation of vocal fatigue.

RESEARCH HYPOTHESIS AND OBJECTIVE

We hypothesized that the severity level of vocal fatigue varies with vocal distance doses the subjects have performed during VLTs. The overall goal of this study is to investigate signs of vocal fatigue using a novel VLT design based on vocal distance dose. The rationale of using vocal distance dose is that the vocal distance dose considers both loudness and pitch in the quantification of voice use, which may be more accurate in voice use control than using loudness and duration control.

METHODS

Subject Data and Experimental Protocol

The protocol included six successive VLT sessions. For each session, cumulative vocal distance was displayed online. In the first session, the subjects were asked to reach a vocal distance of one kilometer within a period of 25 minutes. They were then asked to improve their performance (i.e. reach the same distance over a shorter period) in the following sessions. The subjects' voice qualities and vocal symptoms, such as pitch range, soft phonation quality, and the occurrence of coughing and swallowing, were hypothesized to change as the sessions progressed. Voice was recorded using a condenser microphone calibrated by a Class I sound pressure level meter. Typical fatigue-indicative symptoms such as coughing, swallowing, and voice clearing were manually recorded. Two subjective voice assessment methods (the

CAPE-V and the SAVRa ratings) and acoustic measures were used to find trends of voice quality changes across sessions. A cross-subject comparison on voice quality changes was made in the end.

The human research ethics protocol (A09-M46-11A) was approved by the Institutional Review Board at McGill University. Four female American English speakers were recruited to participate in this experiment. Participant information and experimental conditions are shown in Table 1. All experiments were conducted in a sound proofed voice recording studio.

Table 1: Participant data and experimental conditions

Participant Data				
Subject ID	001	002	003	004
Age	21	25	37	21
Gender	F	F	F	F
Room Temperature	23.6℃	23.7℃	23.7℃	23.7℃
Relative Humidity	39%	27%	33%	29%

The VLT included 6 reading sessions varying from 10 minutes to 25 minutes, and 8 voice assessment sessions during 5-minute breaks. The entire protocol could take up to 4 hours in total. The reading material was the novel "Harry Potter and the Sorcerer's Stone". Participants were asked to hydrate their throats with 100 ml water immediately before each voice assessment session. Drinking was not allowed during VTL sessions

Instrumentation

The VLTs were monitored using a short-time (20-second long) distance dose calculator and a cumulative distance dose calculator in LabView that are shown in Figure 2. When cumulative distance dose reached a fixed value (800 meters for the 1st subject and 1000 meters for the other three subjects), the circular progress LED indicator turned
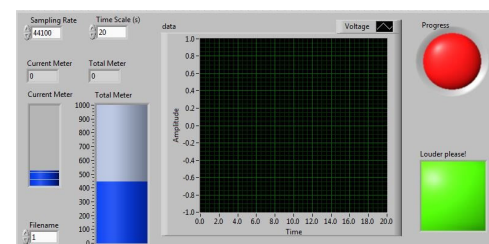

Fig. 2: Virtual vocal distance dose monitor.

from green to red, indicating the completion of one VLT session. The square-shaped loudness-monitoring LED indicator turned from green to red if the participant's previous 20 s distance dose did not reach a threshold value that was set to 15 meters for the first VLT and was increased by 3 meters session by session. Participants were asked to keep an eye on the two indicators while performing the VLTs. The online performance display kept motivating participants to read intensely and louder from one session to the next.

RESULTS

Participants' fatigue-indicative symptoms during the VLTs are shown in Figure 3. Five symptoms, including coughing, swallowing, voice clearing, voice activation, and voice cut out were noted during the VLTs. The SAVRa scores and CAPE-V ratings are shown in Figure 4. In Figure 3, the left y-axis shows symptom counts. The right y-axis shows the time the subjects spent on each session. The gliding pitch results are shown in Figure 5.

DISCUSSION AND CONCLUSIONS

Figure 3 shows visible trends of increased vocal fatigue induced by the vocal loading tasks when the target distance dose was 6000 meters. No obvious fatigue-indicative symptoms were observed for the 4800 meters VLTs. For the third and fourth subjects, increasing loudness (indicated by decreasing session duration) resulted in significantly increased counts of swallowing and coughs. The SAVRa scores on the third row of Figure 4 shows that both the effort to speak loudly and laryngeal discomfort levels increased as the VLTs proceeded and decreased after 10 minutes of vocal rest. No specific trends were found based on CAPE-V ratings. Figure 5 showed a trend of fluctuating maximal pitches for the 2nd, 3rd, and 4th subjects. The maximal pitches noticeably decreased after 3000 meters, and then slightly increased as the VLTs continued. The final 10 minutes rest period did not help increase the pitch range. For the 1st subject, the pitch maximum increased to a peak after 4000 meters
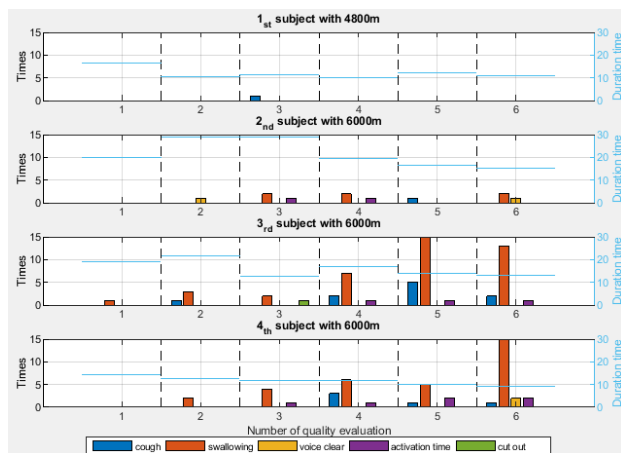

Fig. 3: Results of voice-related measures for individual participants.

and then declined until the end of the 10-minutes rest. fluctuation of the 1st participant's pitch range across sessions is smaller than that of the other three participants. In summary, the different target vocal distance doses (4800 vs 6000 m) induced different levels of vocal quality difference and pitch gliding performance. The maximum pitch was found to be sensitive to vocal fatigue.
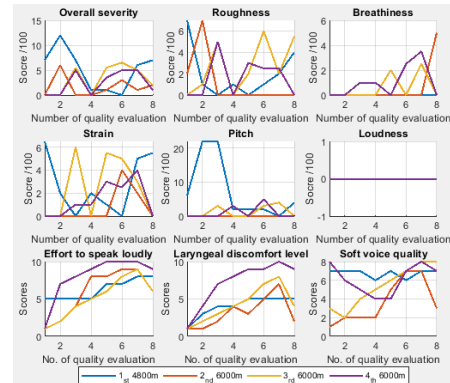
Fig. 4: Results of CAPEV and SAVRa ratings.
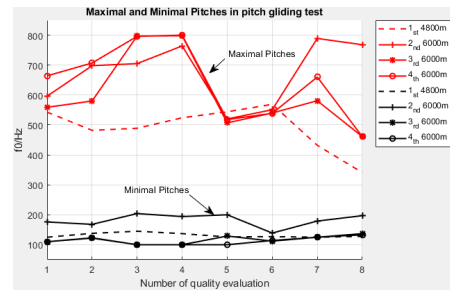
Fig. 5: Maximal and minimal pitches in pitch gliding test.

REFERENCES

[1] N. V. Welham and M. A. Maclagan, "Vocal Fatigue: Current KNowledge & Future Directions," Journal of Voice, vol. 17, no. 1, pp. 21-30, 2003.
[2] R. Brighton Fujiki and M. Preeti Sivasankar, "A Review of Vocal Loading Tasks in the Voice Literature," Journal of Voice, vol. 31, no. 3, pp. 338.e33-388.e39, 2017.
[3] A. Chang and M. P. Karnell, "Perceived Phonatory Effort & Phonation Threshold Pressure Across Prolonged Voice Loading Task: A Study of Vocal Fatigue," Journal of Voice, vol. 18, no. 4, pp. 454-466, 2004.
[4] M. McHenry, J. Evans, and E. Powitsky, "Vocal Assessment Before, After, and the Day After Opera Performance," Journal of Voice, vol. 30, no. 2, pp. 186-191, 2015.
[5] I. R. Titze and E. J. Hunter, "Comparison of Vocal Vibration-Dose Measures for Potential-Damage Risk Criteria," Journal of Voice, vol. 58, pp. 1425-1439, 2015.
[6] Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., and Glaze, L. E., "Establishing Validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V), ". *Am J Speech Lang Pathol*, 20(1), 14-22. doi: 10.1044/1058-0360(2010/09-0105).
[7] Eric J. Hunter, " General statistics of the NCVS self-administered vocal rating (SAVRa)," *The National Center for Voice and Speech Online Technical Memo* N0. 11, Aug. 2011, version 1.1.

AEROACOUSTIC SOUND GENERATION DURING PHONATION – ANALYSIS OF ACOUSTIC ANALOGIES FOR A HYBRID ACOUSTIC PIV APPROACH

Alexander Lodermeyer[1,2], Matthias Tautz[1], Eman Bagheri[1], Stefan Becker[1], Michael Döllinger[3], Stefan Kniesburges[3]

[1] Institute of Process Machinery and Systems Engineering, Friedrich-Alexander University Erlangen-Nuremberg, Germany
[2] Erlangen Graduate School in Advanced Optical Technologies, Friedrich-Alexander University Erlangen-Nuremberg, Germany
[3] Division for Phoniatrics and Pediatric Audiology, Department of Otorhinolaryngology, Head and Neck Surgery, Medical School, Friedrich-Alexander University Erlangen-Nuremberg, Germany

Keywords: **Phonation; Aeroacoustic analysis; Sound generation; Particle image velocimetry**

INTRODUCTION

The investigation of the sound generation mechanisms for healthy and pathological voice production is of scientific and clinical interest. Thereby, the unsteady flow field is a key factor for sound source analysis. In this context, particle image velocimetry (PIV) has been established in voice research to acquire instantaneous 2D velocity fields. By further processing this data, we gain insight into sound generation and acoustic source localization.

METHODS

By using high-speed PIV in a mid-coronal plane within our synthetic larynx model with self-oscillating vocal folds, the supraglottal flow field is resolved up to a sampling rate of 10 kHz. In order to numerically analyze the aeroacoustic sound generation, we applied finite element method (FEM) based simulations (Lighthill, Vortex Sound) and surface integration based (Curle) acoustic analogies. After source term calculation for the earlier, the inhomogeneous acoustic wave equation was solved with the finite element method to obtain the acoustic radiation into the far field.

RESULTS & DISCUSSION

By applying this hybrid approach of high-speed PIV measurements and acoustic simulations, we successfully computed the sound pressure in the far field shown in Fig. 1. Several characteristics in the sound spectra are reproduced in the simulations when compared to the experimental microphone measurements. Both the FEM based and the surface integration based analogies reproduce the broadband shape as well as the tonal components. While all analogies show the tonal components in the spectrum, the Curle approach comprised 18 dB of additional broadband noise. Additionally, reflexions of the vocal tract, which are distinctively visibly above 1 kHz, are not included for the Curle approach. In contrast, the FEM based simulations as well as the microphone measurements include that effect.

During phonation, the highest intensity of tonal component sources was found at the exit of the glottal duct. Broadband sound generation was detected along the shear layers of the glottal jet during the open phase of the glottis.
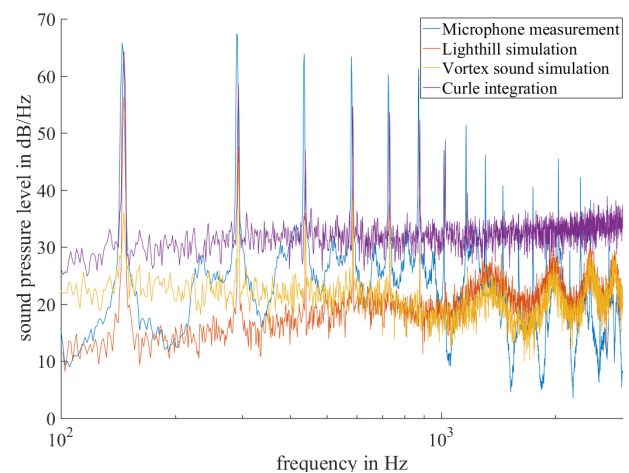
Fig. 1: Resulting spectra of the sound pressure level. Results are shown for the experimental microphone measurement (blue), the FEM simulation based on Lighthill (red) and Vortex sound (yellow), and the simulation by integration as proposed by Curle (purple).

CONCLUSION

Although our technique is restricted to a 2D flow field, we verified the general feasibility of our approach in voice research and determined basic aeroacoustic sound generation mechanisms. Additionally, we found less noise in the FEM-based simulations when compared to Curle's approach.

In a subsequent step, our technique may lead to a better understanding of the sources of voice disorders.

REFERENCES

[1] Becker S, Kniesburges S, Müller S, Delgado A, Link G, Kaltenbacher M (2009) Flow-structure–acoustic interaction in a human voice model. J Acoust Soc Am 125(3):1351–1361.

[2] Curle N (1955) The influence of solid bodies upon aerodynamic sound. Proc R Soc Lond 231(1187):505–514.

[3] Haigermoser C (2009) Application of an acoustic analogy to piv data from rectangular cavity flow. Exp Fluids 47:145–157.

[4] Lodermeyer A, Becker S, Döllinger M, Kniesburges S (2015) Phase-locked flow field analysis in a synthetic human larynx model. Exp Fluids 56(77):1–13.

[5] Lodermeyer A, Tautz, M, Becker S, Döllinger M, Birk, V, Kniesburges S (2018) Aeroacoustic analysis of the human phonation process based on a hybrid acoustic PIV approach. Exp Fluids 59(13):1–15.

[6] Oren L, Koshla S, Gutmark E (2014) Intraglottal geometry and velocity measurements in canine larynges. J Acoust Soc Am 135(1):380–388.

VALIDATION OF A SET OF RECURRENCE QUANTIFICATION MEASUREMENTS IN DISCRIMINATION OF NORMAL AND DEVIATED SYNTHESIZED VOICES

Leonardo Lopes[1], Mara Behlau[.2]

[1] Speech-Language Pathology and Audiology Deapartment, Federal University of Paraíba, João Pessoa, Paraíba, Brazil
[2] Centro de Estudos da Voz (CEV), Federal University of São Paulo, São Paulo, São Paulo, Brazil

***Keywords*: Voice; Vocal quality; Acoustic analysis; Nonlinear analysis**

INTRODUCTION

Acoustic analysis based on nonlinear dynamics has been shown to be a complementary alternative for voice evaluation. This model has been applied [1] in the study of sound signal alterations associated with laryngeal lesions, to differentiate healthy from deviated voices, to classify voices with different degrees of deviation and to evaluate voice disorder treatments.

However, traditional nonlinear measures (such as a correlation dimension D2 and the maximal Lyapunov exponent) were unable to reliably evaluate type 4 signals. This induced the authors to investigate new nonlinear dynamic analysis models of vocal production that can be applied to analyze more deviant signals and provide a more reliable characterization of them [2]. In this context, recurrence quantification analysis (RQA) is a nonlinear dynamic analysis technique for the extraction of features of a nonstationary time series [3]. RQA can be used reliably with nondeterministic, nonstationary, chaotic, and short-duration systems.

In the study of voice, RQMs have an accuracy of >90% in discriminating between patients with and without laryngeal disorders [5]. RQMs had an 85% performance rate in classifying children with healthy and deviant voices (identified by auditory-perceptual analysis) [6]. The results obtained in the cited studies and the extensive application of RQMs in other biological scenarios showed the possibility of further studies in which such measures are used to analyze voice signals.

RQMs demonstrated that are promising for vocal signal evaluation. However, to validate these measures it is necessary to analyze signals in controlled situations, as with the synthesized voices. Therefore, the objective of this research was to identify the set of recurrence quantification measurements with greater accuracy to classify normal and deviant synthesized vocal signals.

METHODS

Samples

A database of 871 synthesized voices generated by the SimuVox synthesizer [7] was used, corresponding to the sustained vowel /ɛ/, including 426 (48.8%) female and 445 (51.2%) male voice signals, with different combinations of the aforementioned acoustic parameters. The signals produced by this synthesizer were previously evaluated for quality and naturalness and were considered to be excellent representations of similar healthy and deviated human vocal production that considered the different types and degrees of deviation. The samples show variations in the vocal deviations of roughness, breathiness and strain.

Instrumentation and Measurement

A visual analogue scale (VAS) with a metric from 0 mm (no voice deviation) to 100 mm (severe deviation) was used for the auditory-perceptual analysis of the voice to evaluate the overall severity of vocal deviation (General grade—GG) [8]. This evaluation was conducted by three speech therapists. Thus, the cutoff values of the VAS GG [8] were used to identify the presence or absence of vocal deviation and subsequently allocate individuals into the appropriate group (with or without a vocal deviation). A total of 47 samples had NVVQ (GG ≤ 35.5 mm) and 824 had vocal deviation (GG ≥ 35.6 mm). Among the samples with vocal deviation, 253 had a mild to moderate degree of deviation (35.6 ≤ GG ≤ 50.5 mm), 550 had a moderate degree of deviation (50.6 ≤ GG ≤ 90.5 mm), and 21 had a severe degree of deviation (GG > 90.5 mm).

For nonlinear acoustic evaluation, we obtained 12 RQMs related to the generation of the three parameters of RPs were observed: determinism (DET), type 1 (T^1) and type 2 (T^2) recurrence time, transitivity (TRANS), average length of diagonal lines (L_{avg}), maximum length of diagonal lines (L_{max}), Shannon's entropy (ENTR), divergence (DIV), average length of vertical structures (also known as trapping time, TT), maximum length of vertical structures (V_{max}), laminarity (LAM), and recurrence period density entropy (RPDE).

Analysis

We used logistic regression to determine a set of measures able to explain the presence of vocal deviation. ROC curve was used to determine the values of accuracy, specificity and sensibility. For complementary analysis we used: positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR[+]) and negative likelihood ratio (LR[-]).

RESULTS AND DISCUSSION

From the adjusted logistic regression model, seven recurrence quantification measurements explained the presence / absence of vocal deviation in the signals studied, including DET, Lmed, ENTR, LAM, TT, RPDE and TRANS (Table 1). The values of the odds ratios showed that the increase in the measures of DET, Lmed and TRANS decreases the chance of vocal deviation, whereas the increase of the values of the ENTR, LAM, TT and RPDE increases the possibility of vocal deviation. The TRANS and RPDE measures showed to be more sensitive to the slight changes in the vocal signal.

Table 1: **Parameter estimates, standard errors, p-values, Odds ratios and 95% confidence intervals of the logistic regression model.**

Variable	Estimate	Standard error	p-value	Odds ratio	LI	LS
Constante	11,8959	14,949	0,4262	-	-	-
DET	-0,8854	0,248	0,0004	2,4240	1,541	4,144
Lmed	-0,6401	0,246	0,0094	0,5272	0,311	0,841
ENTR	8,1121	3,316	0,0144	0,0003	2,61 x 1	0,135
LAM	0,3277	0,166	0,0485	0,7206	0,481	0,951
TT	6,8947	1,848	0,0002	987,0296	43,221	$6,65 \times 10^{+4}$
RPDE	40,4648	14,797	0,0062	0,0175	0,001	0,318
TRANS	-71,9962	19,649	0,0002	0,0007	1,59 x 1	0,035

A set of seven measures selected by the model presents high values of accuracy (0.95), sensitivity (0.95) and specificity (0.96), and an area under the curve of 0.96 (Figure 1).

The performance results in the classification performed through the model are summarized in Table 2. The model presented PPV equal to 99.8%, erring only two classifications. On the other hand, 36 False Negative were counted, that is, individuals with vocal deviation were classified without deviation, as predicted by the model. The values of the LR^+ and LR^- demonstrate the high probability of correctness of the assessment of vocal deviation with the use of this set of measures.
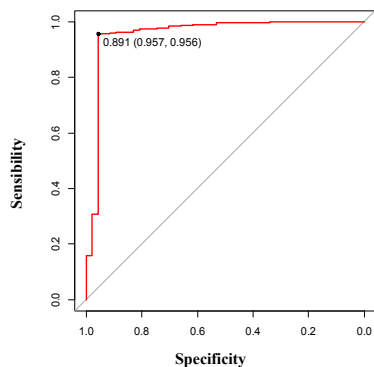

Figure 1: ROC curve for logistic regression model - recurrence variables.

Table 2: Classification of the number of signals with and without vocal deviation according to the logistic regression model.

Predicted by model		Vocal deviation	
		Yes	No
Vocal deviation	Yes	788	2
	No	36	45
Total		824	47

PVV: 788/(788+2) = 99.8%

NPV: 45/(36+45) = 55.5%

$LR^+ = 21.8\%$

$LR^- = 0.04\%$

In general, the model adequacy measure (pseudo-R^2) was 0.63, indicating an excellent performance to explain the presence / absence of vocal deviation.

CONCLUSION

The set of seven RQMs present excellent performance in the classification of normal and deviant synthesized signals.

REFERENCES

[1] Cheng Y, Li ZQ, Huang JZ, *et al*. Combination of autologous fascia lata and fat injection into the vocal fold via the cricothyroid gap for unilateral vocal fold paralysis. Arch. Otolaryngol. Head Neck Surg. 2009; 135:759–763.

[2] Calawerts WM, Lin L, Sprott JC, Jiang JJ. Using rate of divergence as an objective measure to differentiate between voice signal types based on the amount of disorder in the signal. J. Voice. 2017;31(1):16–23.

[4] Marwan N, Webber CL. Mathematical and Computational Foundations of Recurrence Quantifications. In: Webber Jr C, Marwan N, eds. *Recurrence Quantification Analysis. Understanding Complex Systems*. Cham: Springer; 2015:3–43.

[5] Costa WCA. Dynamic non-linear analysis of voice signals for detection of laryngeal disorders. [Thesis]. Federal University of Campina Grande; 2012.

[6] Lopes LW, Costa SL, Costa WC, Correia SE, Vieira VJ. Acoustic assessment of the voices of children using nonlinear analysis: proposal for assessment and vocal monitoring. J. Voice. 2014;28(5):565–573.

[7] Lucero JC, Schoentgen J, Behlau M. Physics-based synthesis of disordered voices. Proceedings of Interspeech 14th Annual Conference of the International Speech Communication Association. 2013; 25-29.

[8] Yamasaki R, Madazio G, Leão SHS, Padovani M, Azevedo R, Behlau M. Auditory-perceptual evaluation of normal and dysphonic voices using the voice deviation scale. J Voice. 2017; 31(1):67-71.

AMNIOTIC FLUID ASSISTS IN SEPARATION OF MURINE VOCAL FOLDS

Vlasta Lungova[1], Tadeas Lunga[1, 2], Susan L. Thibeault[1]

[1] Department of Surgery, University of Wisconsin-Madison, Madison, Wisconsin, United States
[2] Faculty of Medicine and Dentistry, Palacky University Olomouc, Hnevotinska 3, Olomouc, Czech Republic

Keywords: **Amniotic fluid, Epithelial Lamina, Remodeling**

INTRODUCTION

The main aim of this study was to find out the effect of the drainage of the amniotic fluid on vocal fold recanalization.

METHODS

Five pregnant females at day 15.5 gestation [embryonic (E) day 15.5], underwent surgery. The peritoneum was open and embryos were removed. Depending on the size of the litter, from two to four embryos underwent a puncture of the amniotic sac with a 21G needle (Fig. 1). Remaining embryos were left as controls. After perforation embryos were returned to their initial position. The peritoneal cavity was flushed with saline and closed with sutures, the skin was closed with clips. Tissue collection was performed day 1 and 2 post-surgery.
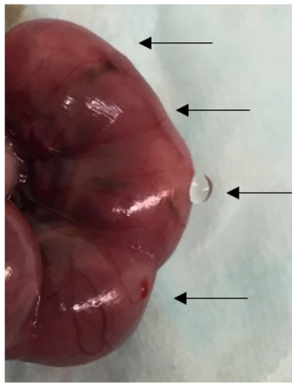
Fig. 1: Drainage of amniotic fluid.

RESULTS

In control and drained embryos, vocal folds (VF) fully separate by E17.5. However, while in controls VFs separate gradually, at E16.5 only a small portion of the epithelial lamina is still apparent, drained embryos show a delay in the VF separation. At E16.5 the epithelial lamina is well visible. It rapidly disintegrates at E17.5, as the flow of the amniotic fluid is restored. This "forced" VF separation affected the VF morphology. They formed polyp-like structures with mesenchymal core (Fig. 2) and were covered with cytokeratin (K) 8 positive epithelial cells. Staining for a basal cell marker p63 revealed that basal cells covering the polyp lost p63 expression, suggesting that the epithelium could undergo remodeling. Moreover, the epithelium lining VFs had signs of epithelial damage. The suprabasal cell layers peeled off, so the basal p63 cells flattened to protect the lamina propria and induce regeneration.
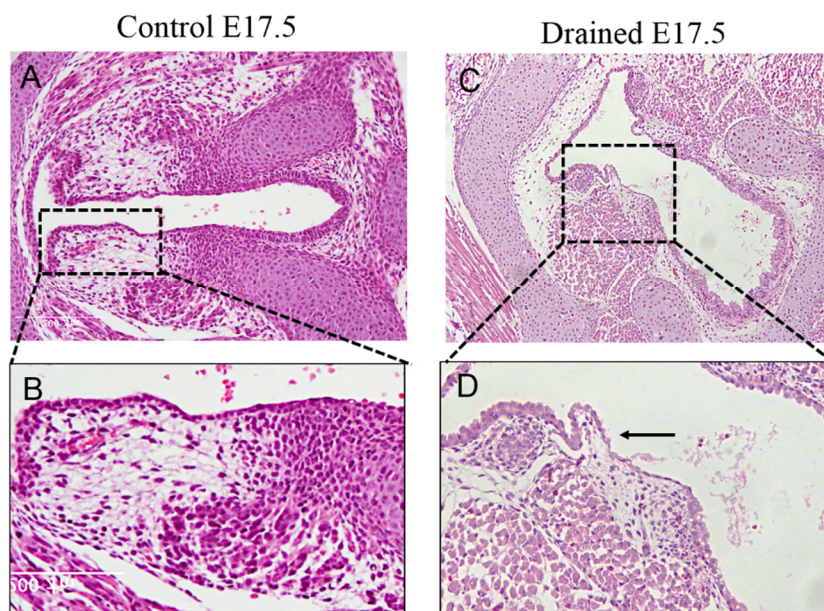
Fig. 2: VF morphology in control (A, B) and drained embryos (C, D).

CONCLUSION

Our results show that amniotic fluid assists in VF separation. Moreover, exposure of damaged epithelial cells to mechanical forces can lead to polyp formation in VFs.

EFFECTS OF A MASK RIM LEAK ON AC AIRFLOW MEASUREMENT

Nicholas May[1], Ronald Scherer[1]

[1] Department of Communication Sciences and Disorders, Bowling Green State University, Bowling Green, Ohio, U.S.A.

Keywords: Pneumotachograph; Pneumotach; Mask Leak; Airflow Measurement

INTRODUCTION

Measurement of airflows during speech often utilize a pneumotachographic mask system. Measurement accuracy depends upon the assumption of a complete seal of the mask rim to the face. The primary aim of this study was to determine the relationship among mask rim leak cross-sectional area, alternating current (AC) cycle frequency, AC peak-to-peak airflow amplitude, and magnitude of upstream direct current (DC) airflow (independent variables) on AC peak-to-peak airflow through the mask (dependent variable) in order to determine the magnitude of AC airflow measurement error. This study extends previous empirical bench research [3] to AC airflow conditions such as vibrato or tremor.

METHODS

Bench Set-Up

Steady airflow was supplied by a wet-dry vac and regulated by adjusting variable transformer voltage and the use of flow valves (**Figure 1**). A deformable rubber tube was placed in series with the PVC pipe system to allow modulation of airflow (i.e., creation of various AC rates and extents) via manual compression. A calibrated Validyne MP45-16-87 pressure transducer was connected to a calibrated Rudolph 37888 pneumotachograph. Thus, it was possible to calculate airflow for any voltage from the pneumotach-transducer system when using the mask system attached downstream of the pneumotach. The Glottal Enterprises (GE) flow mask system (MSIF-2), with its calibration mold and mask with calibrated PTW-1 pressure transducer, was attached downstream of the pneumotach to capture the pressure drop across the mask. The outputs of the GE MSIF-2 system and the Validyne carrier demodulator were connected to a DATAQ analog-to-digital converter, which was connected to a computer for display and recording via DI-2108 software. The modulating airflow signal was created by compression by hand of the deformable tube upstream of the pneumotach.

Mask rim leaks were simulated by *c.* 45 mm long brass tubes with either a rectangular or quasi-elliptical cross-sectional area. The leak tubes were placed between the mask and the mold with pliable putty to prevent unmeasured leaks (**Figure 2**). Leaks were "added" or "subtracted" by manual occlusion or release, respectively, of the end of the tubes.
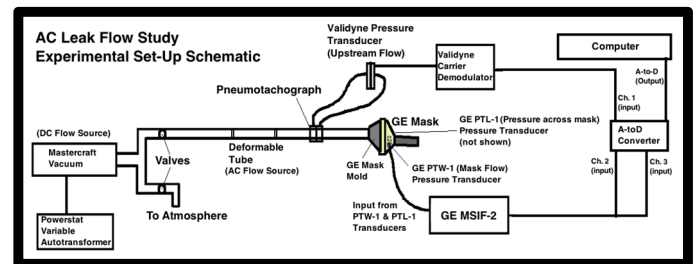
Fig. 1: Schematic of the experimental bench set-up.

Fig. 2: Picture of leak tube 6A in place between the mask and the mold.

Independent variables

Three leak areas (0.34 cm^2, 0.73 cm^2, and 1.30 cm^2), two steady upstream airflows (a lower upstream airflow of *c.* 100 to 165 cm^3/s and a higher upstream airflow of *c.* 330 cm^3/s), two AC airflow frequencies (3 Hz and 6 Hz), and a range of peak-to-peak airflow amplitude variations (*c.* 50 to 200 cm^3/s) were used.

Analysis

Signals recorded with DATAQ DI-2108 software were exported into Sigplot (MATLAB-based custom software) for transformation to physical units and for zeroing of the airflow baseline. Sigplot outputs were then exported to Microsoft Excel for display. In Excel, a moving average was generated to approximate the DC airflow level during the AC airflow generation (**Figure 3**). Representative samples were extracted from Excel and imported into a custom MATLAB peak-picking program. Peaks and troughs for the leak and no leak conditions were selected, peaks and troughs in each condition were averaged, and peak-to-peak values were calculated by subtracting the values of the mean troughs from the values of the mean peaks. Paired samples t-tests were performed to determine whether mean peak-to-peak values were statistically different from each other ($\alpha = .05$). Percent peak-to-peak reductions from the no leak to leak condition were also generated.
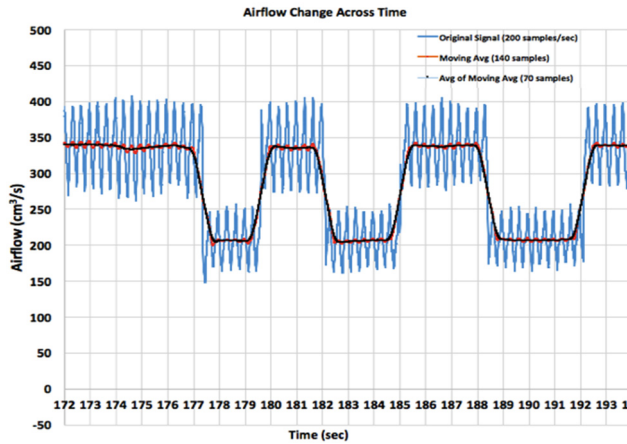
Fig. 3: Microsoft Excel display of an AC airflow condition. Leak tube #9 (area: 0.73 cm^2), upstream airflow: 330 cm^3/s, AC rate: 3 Hz, AC pk-to-pk: 100 cm^3/s. The black line represents the approximate DC airflow level given no leak (top portion) and a leak (shifted down portion).

RESULTS

When comparing leak to no leak conditions, 11/12 paired t-tests reached statistical significance ($\alpha = .05$); 12/12 leak conditions had lower mean pk-to-pk values than the no leak conditions, but one condition failed to reach statistical significance. A lower pk-to-pk value means that there was a reduction in the AC airflow signal measured by the mask when there was a rim leak, indicating that the mask system reported incorrect airflow values because of the mask rim leak. When comparing the pk-to-pk values for the 3 Hz and 6 Hz AC frequency leak conditions (i.e., with all other variables held constant), the 6 Hz condition resulted in larger pk-to-pk values than the 3 Hz condition in 10/12 comparisons; however, only 7/12 comparisons reached statistical significance ($\alpha = .05$). When comparing the pk-to-pk values for the low upstream airflows (100 to 165 cm^3/s) to the high upstream airflow (*c.* 330 cm^3/s) conditions (i.e., with all other variables held constant), the results were equivocal with higher pk-to-pk values found in the lower upstream airflow leak condition in 7/12 cases and higher pk-to-pk values in 5/12 cases for the higher upstream airflow condition. These inconsistent results are probably caused by the use of separate airflow trials for low and high upstream airflows and the inherent variability in the manual pumping between trials. In addition to the paired t-test comparisons, mean percent reduction in pk-to-pk airflow caused by the presence of a leak were generated (**Table 1**). This table was generated by dividing the pk-to-pk value in the leak condition by the pk-to-pk value in the no leak condition for all conditions, subtracting each of these decimal values from 1 (yielding a % reduction in pk-to-pk airflow given a leak), and averaging the values for each factor (e.g., for 3Hz, for 6 Hz).

Table 1: Mean % reduction in pk-to-pk airflow given a leak.

	Mean % reduction in pk-to-pk airflow given a leak	SD
3 Hz	**-24.5%**	**11.1%**
6 Hz	**-28.5%**	**10.2%**
50 pk-to-pk	**-24.4%**	**11.1%**
200 pk-to-pk	**-28.6%**	**10.1%**
100 (or 165) cm^3/s upstream airflow	**-29%**	**9%**
330 cm^3/s upstream airflow	**-23.9%**	**11.9%**
Tube 6A (0.34 cm^2)	**-15.9%**	**9.1%**
Tube 9 (0.73 cm^2)	**-28.1%**	**7.6%**
Tube 8+9 (1.30 cm^2)	**-35.3%**	**3.2%**

DISCUSSION AND CONCLUSION

Generally, the larger the leak cross-sectional area, the larger the proportional reduction in both the DC airflow level [3] and in the pk-to-pk AC airflow measured at the mask. Previous research has indicated that a mask rim leak may act as a high-pass filter [2] or that a "relatively small" mask leak should have a negligible effect on measurement of airflow amplitude due to the "high acoustic impedance of the leakage path" [1]. The results of this study indicate that mask rim leaks of 0.34, 0.73, and 1.30 cm^2 result in approximately 16%, 28%, and 35% reductions in pk-to-pk airflow amplitude, respectively, suggesting that smaller leaks result in smaller proportional losses. These findings should be interpreted with caution due to the un-ecological nature of the simulated leaks (i.e, the tubes are longer and of different shape than actual leaks at the rim-face interface). In summary, the primary findings of this project are the following:

1. A mask rim leak (0.34 - 1.30 cm^2) results in a substantial reduction in pk-to-pk airflow amplitude (16% - 35%) as measured by a GE aerodynamic mask system.

2. Larger leaks result in larger reductions in pk-to-pk airflow amplitude (i.e., larger reductions in AC airflow).

3. Given a leak, a larger reduction in pk-to-pk airflow amplitude (-28.5%) was found for the 6 Hz conditions than for the 3 Hz conditions (-24.5%) indicating that a mask-rim leak may act as a low-pass filter.

4. Given a leak, a larger reduction in pk-to-pk airflow amplitude (-28.6%) was found for the large (200 cm^3/s) pk-to-pk conditions than for the small (50-100 cm^3/s) pk-to-pk conditions (-24.4%) indicating that a mask-rim leak may be more resistive to large pk-to-pk amplitude variations.

REFERENCES

[1] Higgins M.B. & Saxman, J.H. (1991). A comparison of selected phonatory behaviors of healthy aged and young adults. *J Speech Lang Hear Res*, 34, 1000–1010.
[2] Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech, Language, and Hearing Research*, *38*(6), 1212-1223.
[3] May, N. A., & Scherer, R. C. (online 2017). Airflow error measurement due to pneumotachograph mask rim leaks. *Journal of Voice*.

DESIGN AND EVALUATION OF PATIENT-SPECIFIC LARYNGEAL IMPLANTS

Michael J. McPhail[1], Justin Hintze[1], Cheryl E. Myers[1], David G. Lott[1,2]

[1]Head and Neck Regeneration Program, Center for Regenerative Medicine, Mayo Clinic, Scottsdale, AZ, USA
[2]Department of Otorhinolaryngology, Mayo Clinic, Scottsdale, AZ, USA

Keywords: **e.g. Laryngeal implant; Patient-specific; 3D Printing; Modeling**

Abstract:

Porous polyethylene (Medpor) was shaped with 3D printed molds to produce patient-specific hemi-laryngeal implants.

Objectives: The long-term goal of this work is to restore proper laryngeal function to patients suffering from laryngeal cancers or defects.

Methods: Computed tomography (CT) scans were acquired on patients suffering from laryngeal cancers and defects. Detailed scans were conducted with patients holding their glottis open and images were segmented with 3D Slicer software. Hemi-larynx implants were designed by mirroring the unaffected side of the larynx. Meshmixer software was used to trim the segmented hemi-larynx and medialize the vocal fold. A 3D model of the mold was designed in Autodesk 123D design software and 3D printed with a stereolithography printer. The mold was then used to shape the porous polyethylene implant in a patient-specific hemi-larynx three-dimensional shape. The effects of the molding process on the porous polyethylene mechanical properties, pore size and distribution were characterized.

Results: The molding procedure was found to have minimal effects on the structure of the porous polyethylene. The molded implant reliably reproduced and retained the geometry of the design. The implant molding procedure was able to be conducted in a sterile manner in an operating room.

Conclusions: The described technique produces a rigid, patient-specific hemi-laryngeal implant out of a material commonly used in head and neck surgery. The implant can be designed, from CT scan to surgery, in under a week.

IN VIVO PROBE FOR TRACKING INTRAGLOTTAL PRESSURE, VOCAL FOLD COLLISION, AND SUBGLOTTAL PRESSURE DURING PHONATION

Daryush Mehta[1,2,3], James Kobler[1,2], Steven Zeitels[1,2], Robert Hillman[1,2,3]

[1] Center for Laryngeal Surgery & Voice Rehabilitation, Massachusetts General Hospital, Boston, MA, USA
[2] Harvard Medical School, Boston, MA
[3] MGH Institute of Health Professions, Boston, MA, USA

***Keywords*: Subglottal Pressure; Intraglottal Pressure; Vocal Fold Collision, Phonotrauma**

INTRODUCTION

A critical element in the development of ambulatory vocal dose measures is the characterization of vocal fold collision [1], which is widely considered to be a primary etiological factor of phonotrauma. In their original formulation, vocal dose measures were designed to indirectly qualify the accumulated effects of rapid acceleration and deceleration of vocal fold tissue (distance dose) and the breaking of molecular bonds at the cellular level due to thermal agitation (energy dissipation dose) [2]. The inclusion of collision/impact stress is important in any quantification or modeling since phonotrauma is thought to arise due to repetitive stress on the mid-membranous portion of the vocal folds [3, 4].

Since measuring collision pressures during bilateral vocal fold phonation has proven to be challenging [5, 6], the goal of the current line of study is to recruit patients with a unilateral cordectomy as a phonatory model in which a pressure sensor rests against the non-vibrating vocal fold to yield stable and reliable collision pressure measurements of the medial surface of the contralateral vocal fold. In addition, it is possible to directly measure the subglottal pressure, which is a primary factor in controlling voice onset, offset, and intensity, and necessary to accurately interpret vocal fold collision data. This paper reports on the development of a new in vivo probe for the simultaneous measurement of intraglottal pressure (collision pressure during contact phase) and subglottal pressure. Initial results are shown from benchtop testing using an airflow-driven excised larynx model.

METHODS

Instrumentation

Simultaneous measurement of intraglottal and subglottal pressures was accomplished using two miniature pressure-sensitive sensors originally designed for heart valve and arterial pressure measurement for cardiovascular function assessment (Mikro-Cath Pressure Catheter, Millar, Inc., Houston, TX). These sensors are 1.17 mm in diameter and were placed inline at the distal end of a modified rigid probe typically used for outpatient vocal fold injections. The sensors were positioned approximately 4 cm apart such that the inferior sensor would measure subglottal pressure and the superior sensor would measure
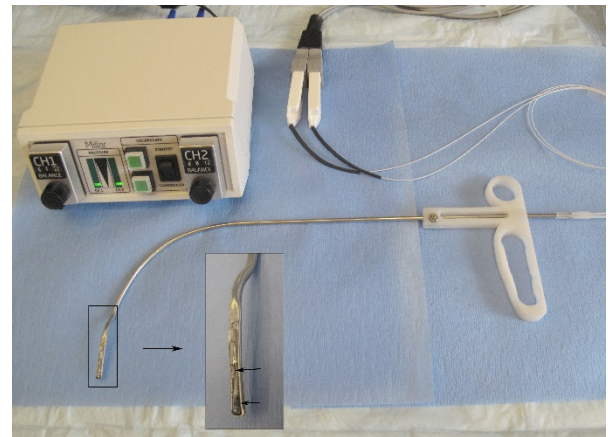

Fig. 1: In vivo probe with dual pressure sensors at the tip (arrows in inset) for simultaneously measuring intraglottal and subglottal pressure during phonation.

intraglottal/collision pressure during phonation. Extensive testing has stressed the importance of having a flat surface for the vocal fold to come into contact with to accurately calibrate collision pressure to sensor output.

The dual-sensor configuration enables the simultaneous capture of intraglottal pressure and subglottal pressure to aid in the separation of the components of the intraglottal pressure sensor signal into an aerodynamic energy component (during the open phase) and vocal fold impact stress component (during the closed phase). This separation is important for modeling the radiation/transfer of energy of these two components.

Excised larynx experiment

A preliminary experiment was conducted using an excised human larynx to simulate unilateral vocal fold phonation using airflow-driven oscillation at varying intensity levels. A whole-mount preparation of the excised larynx was prepared by dissecting away supraglottal structures (hyoid, etc.), suturing the ventricular folds, and leveling the thyroid cartilage to expose the true vocal folds superiorly. Inferiorly, the trachea is cut to a length of approximately 5 cm and the specimen mounted on a cylindrical pipe connected to an airflow supply.

Airflow was sent through a ConchaTherm-IV device (Hudson RCI, Research Triangle Park, NC) that warmed
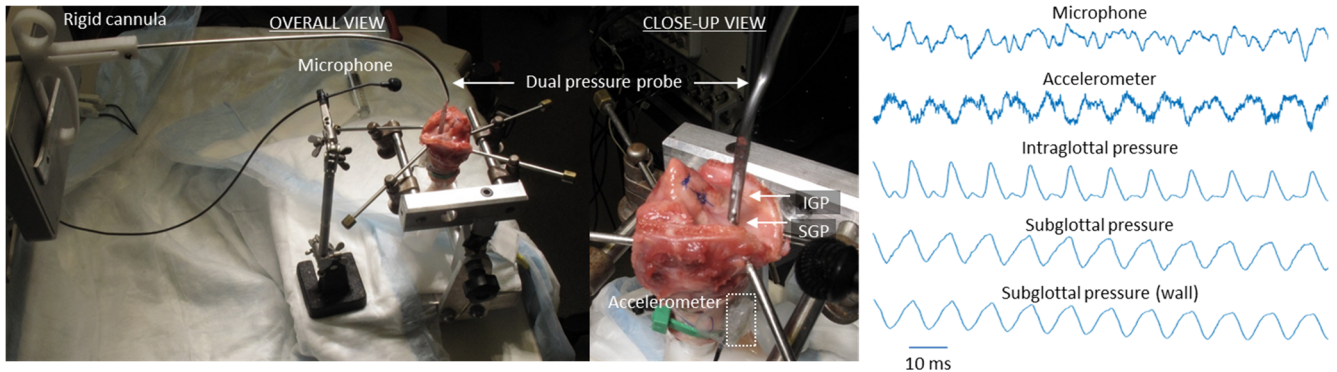
Fig. 2: Human cadaver excised larynx preparation and time-synchronized waveforms from an acoustic microphone, high-bandwidth accelerometer, intraglottal pressure (IGP), and subglottal pressure (SGP).

and humidified the air before directing the stream through the trachea to produce self-driven, self-sustained vocal fold oscillation. A pneumatic pressure gauge regulated the force of the air stream, and a low-bandwidth transducer (Glottal Enterprises, Syracuse, NY) tracked the pressure level approximately 6 cm below the vocal folds (wall pressure).

A typical trial consisted of sustained vocal fold oscillations at a given subglottal pressure and glottal configuration. This setup enables testing measures of vocal fold collision pressures at varying levels of subglottal pressure, fundamental frequency, and vocal fold adduction/abduction. Since ambulatory monitoring of voice use is of primary interest, determining the relationship between vocal fold impact stress/collision pressure and subglottal neck-surface acceleration is crucial. In this benchtop model, a high-bandwidth accelerometer was placed subglottally to mimic the subglottal placement on the anterior neck surface on a human subject.

RESULTS AND DISCUSSION

In agreement with previous in vivo, excised, and numerical studies of vocal fold impact stress, the intraglottal pressure sensor exhibited two components: an intraglottal aerodynamic energy component (during the open phase) and a vocal fold collision peak (during the closed phase). Secondarily, the dual-sensor probe configuration aided in determining correct sensor placement because the signal produced at the intraglottal location was easily differentiated from the subglottal sensor signal.

Future work will investigate the dependence of intraglottal pressure signal properties on changes in sensor motion and position to determine the optimal location for tissue collision tracking and impact of the presence of the sensor on phonation and airflow. Studying both hemilarynx and whole-mount setups will assess the impact of having one versus two vibrating vocal folds during phonation to translate results from unilateral cordectomy subjects to individuals with two vibrating vocal folds.

CONCLUSION

A dual-channel probe was developed to simultaneously measure intraglottal pressure and subglottal pressure signals in a patient with a unilateral cordectomy to gain empirical insight into vocal fold collision and aerodynamic relationships during phonation.

REFERENCES

[1]   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, pp. 1425-1439, 2015.

[2]   I. R. Titze, J. G. Švec, and P. S. Popolo, "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," *Journal of Speech, Language, and Hearing Research,* vol. 46, pp. 919-932, 2003.

[3]   C. Tao, J. J. Jiang, and L. Czerwonka, "Liquid accumulation in vibrating vocal fold tissue: A simplified model based on a fluid-saturated porous solid theory," *Journal of Voice,* vol. 24, pp. 260-269, 2010.

[4]   L. Czerwonka, J. J. Jiang, and C. Tao, "Vocal nodules and edema may be due to vibration-induced rises in capillary pressure," *Laryngoscope,* vol. 118, pp. 748-52, 2008.

[5]   K. Verdolini, M. M. Hess, I. R. Titze, W. Bierhals, and M. Gross, "Investigation of vocal fold impact stress in human subjects," *Journal of Voice,* vol. 13, pp. 184-202, 1999.

[6]   H. E. Gunter, R. D. Howe, S. M. Zeitels, J. B. Kobler, and R. E. Hillman, "Measurement of vocal fold collision forces during phonation: Methods and preliminary data," *Journal of Speech, Language, and Hearing Research,* vol. 48, pp. 567-576, 2005.

GLOTTAL ATTACK TIME IN CONNECTED SPEECH

Maryam Naghibolhosseini[1], Dimitar D Deliyski[1], Stephanie RC Zacharias[2], Alessandro de Alarcon[3], Robert F Orlikoff[4]

[1] Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, U.S.A.
[2] Department of Otolaryngology Head & Neck Surgery, Mayo Clinic, Scottsdale, AZ, U.S.A.
[3] Division of Pediatric Otolaryngology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, U.S.A.
[4] College of Allied Health Sciences, East Carolina University, NC, U.S.A.

Keywords: **High-Speed Videoendoscopy; Glottal Attack Time; Connected Speech; Laryngeal Imaging**

Abstract:

Objectives: The goal of this project is to study the glottal attack time during connected speech using high-speed videoendoscopy (HSV). The glottal attack time measure was computed automatically from the images for each vocalization in connected speech. This measure can be beneficial in studying the pathophysiology of voice disorders.

Methods: A custom-developed HSV system coupled with a flexible fiberoptic nasolaryngoscope was used to record one vocally normal female participant during reading of the "Rainbow Passage". A gradient-based algorithm was developed and used to perform temporal segmentation of the HSV data into the phonatory components of speech. A spatial segmentation algorithm was then developed (based on an active contour modeling approach) to provide an analytical description of the vocal fold edges in the HSV-derived kymograms. The kymogram at the medial section of the vocal folds was used to compute the glottal attack time. The glottal attack time was measured as the elapsed time from the first vocal fold oscillation to the first vocal fold contact in each vocalization.

Results: The result of the temporal and spatial segmentation provided the information regarding the onset and offset of voicing for each vocalized segment of the connected speech along with the extracted edges of the vocal folds. The glottal attack time was then successfully computed for each vocalized segment of the connected speech.

Conclusions: This study demonstrates the feasibility of measurement of glottal attack time in connected speech using HSV. Development of an automated method for glottal attack time measurement can be valuable in the clinical practice to develop new clinical protocols for HSV-based voice assessment of connected speech.

ONSET OF RIGOR MORTIS OF THE VOCAL FOLDS IN AN IN VIVO-EXCISED RABBIT MODEL

Juergen Neubauer [1] , Stephanie Zacharias [2] , David Lott [2]

[1] School of Human Evolution and Social Change, Arizona State University, AZ, USA
[2] Department of Otorhinolaryngology, Mayo Clinic, Scottsdale, AZ, USA

Keywords: Rigor Mortis; Laryngeal Pressure-Flow Relationship; In vivo Rabbit Larynx; Excised Rabbit Larynx;

Abstract:

Objectives: When does rigor mortis in laryngeal muscles occur in a rabbit model? Previous work on rigor mortis onset times in rabbit limb muscles showed large variability ranging from one to nine hours. The rabbit model is commonly used in pre-clinical studies to evaluate the efficacy and safety of clinical voice treatments. Knowledge of rigor mortis of laryngeal muscles is critical for unbiased comparisons of clinical voice treatments.

Methods: In anesthetized New Zealand rabbits, a fully humidified, heated air flow system was connected to the trachea inferior to the cricoid cartilage. All supraglottal structures were removed and arytenoid cartilages were sutured closed. Pressure-flow relationships were obtained repeatedly while the animal was anesthetized and immediately post-euthanasia. Preparation pre/post euthanasia was kept unchanged. Automated air flow ramp (32 seconds duration) was applied every 10 minutes for 4 hours. A LabView control program was used to synchronize and automatize air flow control, data acquisition, and high speed video. The following measurements were recorded: transglottal differential air pressure, subglottal acoustic pressure fluctuations, outside sound pressure, and laryngeal deformation and possibly vibrations with a high speed video.

Results: Animal preparations were challenging because of translucent, soft ventricular fold structures attached to the lateral aspects of the superior vocal fold surfaces. These ventricular folds could vibrate at low onset pressures and occluded the visualization of the true vocal folds. We developed a method to prevent this interference. At the end of the 4 hours of flow ramp measurements, the maximum transglottal pressure significantly increased from about 30 cmH2O to up to 60 cmH2O.

Conclusions: Using linear air flow ramps, the pressure-flow relationships revealed the time course of onset of laryngeal muscle rigor mortis, a critical factor for assessments of efficacy and safety of clinical voice treatments.

A NEW TECHNIQUE FOR MEASURING THE SOURCE WAVEFORM OF (TRACHEO)ESOPHAGEAL SPEAKERS: TWO-POINT TRANSFER MATRIX METHOD

Andressa B. Otto[1], Andrey R. da Silva[1], Ana Carolina Ghirardi[2], Elisa G. Vieira[3]

[1] Department of Mechanical Engineering /Laboratory of Vibration and Acoustics, Federal University of Santa Catarina, Florianópolis, Brazil
[2] Speech-Language Pathology and Audiology Department, Federal University of Santa Catarina, Florianópolis, Brazil
3 Center for Oncological Research of Santa Catarina, Florianópolis, Brazil

Keywords: source waveform, tracheoesophageal speech, voice measuring technique

INTRODUCTION

The assessment of the glottal waveform is an important procedure to obtain key aspects of voice condition, including its spectral content, pulse width and closure, skewness, spectral tilt, pressure threshold for self-sustained oscillation, to name but a few [1].

The majority of the available non-invasive procedures to obtain the glottal waveform are based on the glottal inverse filtering family of techniques, also known as GIF. These techniques use different sorts of algorithms, mostly based on linear prediction, in order to achieve a mathematical representation of the vocal tract in terms of an all-pole filter. This representation is used to filter out the externally measured voice of a patient, thereby obtaining its source signal [2].

A more straightforward technique is the Sondhi tube [4]. The tube consists of a straight pipe segment with one of its extremes terminated by a sound absorbing wedge. The second termination consists of a mouthpiece, through which the person phonates, and her/his voice is recorded by a microphone flush mounted to the tubes wall. The Sondhi tube acts as to eliminate the mouth's radiation impedance and the vocal tract resonances by extending the vocal tract to an acoustically infinite tube.

Although these techniques may provide a good assessment of the source signal of glottal speakers, they may not be adequate for obtaining the source waveform of subjects with esophageal and tracheoesophageal speech for a few reasons. In the case of the GIF family of techniques, the (tracheo)esophageal signal has a relatively high level of aperiodicity [3], which may lead to significant errors on the linear prediction process. Moreover, the idea that the (trecheo)esophageal speech can be dissociated into the source-filter paradigm is a rather uncertain assumption. Likewise, in the case of the Sondhi tube, the low-frequency reflections at the anechoic termination of the pipe and the discontinuity at the interface between the tube and the patient's mouth may add significant resonances to the measured signal.

The objective of this work is to propose a technique for assessing the source waveform of esophageal and tracheoesophageal speakers by using a tube with two microphones. The technique involves determining the subject's vocal tract geometry using a layer-peeling algorithm. Thereafter, the source waveform is obtained from the subject's voice signal by resolving a two-port transfer matrix equation originated from the layer-peeling process.

METHODS

Experimental Setup and Data Acquisition

The experimental setup used in this work consisted of a 1045 mm-long cylindrical duct, with inner diameter of 20 mm, which provided a cut-off frequency for plane wave propagation of 9 kHz. Two ¼" B&K microphones type 4944A were flush-mounted on the duct, as depicted in Fig. 1.
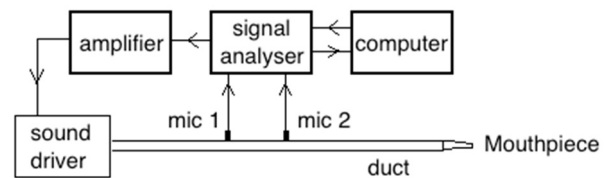
Fig.1: Scheme of the experimental setup.

The acoustic signal generated by a signal analyzer device B&K Pulse type 30-50-A-60 was amplified using a 2735 B&K module and reproduced by a JBL 4400Ti sound driver. The data acquisition involving human subjects is divided into the following steps.

1. The subject is asked to adapt his/her mouth onto the mouthpiece placed at the duct's open end and instructed to mimic the pronunciation of vowel /e/. A broadband signal is reproduced by the driver depicted in Fig. 1, whose duration is five seconds and the system's acoustic response is captured by microphones 1 and 2.

2. The subject's vocal tract geometry is reconstructed using the layer-peeling algorithm [5]. Moreover, the acoustic impedance *Z1* at microphone 1 is determined by wave decomposition using data from microphones 1 and 2 [6]. Based on the vocal tract geometry, a transfer matrix *[M]* representing the vocal tract is constructed [7].

3. The subject is asked to pronounce the vowel /e/ during approximately five seconds. The pressure signal obtained

at microphone 1 is converted to the frequency domain by a discrete Fourier transform, which gives *P1*. Thereafter, the acoustic particle velocity *Us* and the acoustic pressure *Ps* produced by the esophageal segment are obtained by resolving the following system:

$$\begin{bmatrix} Ps \\ Us \end{bmatrix} = [M] \begin{bmatrix} P1 \\ P1/Z1 \end{bmatrix}. \qquad (1)$$

The source waveform in terms of pressure and volume flow are finally obtained by the inverse discrete Fourier transform of *Ps* and *Us*, respectively.

Method Validation

The method for obtaining the source waveform described in the previous Section was validated using an artificial vocal tract coupled to a sound source, whose waveform was known. In this case, instead of measuring a human subject, the artificial vocal tract was connected to the duct's mouthpiece. The data acquisition procedures described in the previous Section were conducted in order indirectly obtain the acoustic pressure *Ps* produced by the source connected to the artificial vocal tract. Figure 2 shows the comparison between the real artificial vocal tract geometry and the reconstructed geometry using the layer-peeling algorithm.
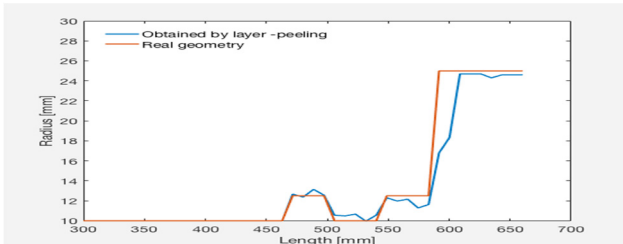

Fig. 2: Comparison between real and reconstructed vocal tract geometric profiles.

Using the reconstructed geometry, the transfer matrix *[M]* was obtained and the acoustic pressure at the source was indirectly assessed using Eq. (1). Figure 3 provides a comparison between the real waveform signal produced by the source and the source waveform obtained indirectly using the methodology proposed in this work.
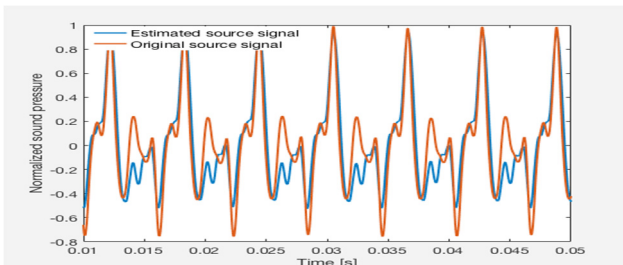

Fig. 3: Comparison between real and reconstructed source waveform.

RESULTS

The method described above was used to measure the source waveform of five tracheoesophageal speakers at the Center for Oncological Research of Santa Catarina. Figure 4 provides the source waveforms obtained for a single speaker using three different techniques, including the Sondhi tube, GIF and the transfer matrix method.
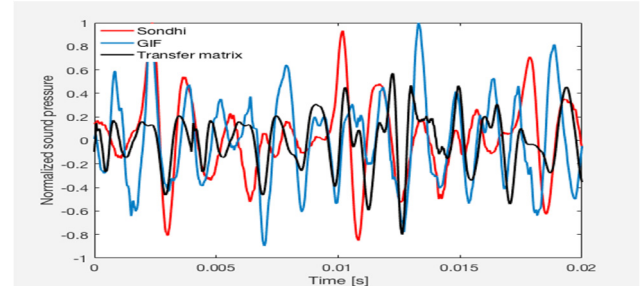

Fig. 4: Comparison between source waveforms obtained with three different techniques.

CONCLUSION

The results depicted in Fig. 4 obtained with the GIF and Sondhi techniques diverge significantly from that obtained with the transfer matrix method. This suggests that the classical techniques may lead to considerable errors when used to estimate the source waveform of (trachea)esophageal speakers. In addition to its simplicity, the transfer matrix technique can be further used to estimate the length and geometry of the vocal tract of tracheoesophageal speakers.

REFERENCES

[1] Monsen RB, Engebretson AM. Study of variations in the male and female glottal wave. *The Journal of the Acoustical Society of America*, 1977, Vol. 63. No. 4, 981-993.

[2] Alku P Glottal inverse filtering analysis of human voice production – A review of estimation and parameterization methods of the glottal excitation and their applications. *Official Journal of the Indian Academy of Sciences*. Vol. 36. Part 5. 2011, 623-650.

[3] QI Y, Weinberg B Characteristics of voicing source waveforms produced by esophageal and tracheoesophageal speakers. *Journal of Speech and Hearing Research*, 1995, 536-548.

[4] Sondhi M M Measurement of the glottal waveform. *The Journal of the Acoustical Society of America*, 1975, Vol. 57. No. 1, 228-232.

[5] Gray C D Acoustic pulse reflectometry for measurement of the vocal tract with application in voice synthesis. PhD thesis, University of Edinburgh, 2005, p.272.

[6] International Standard ISO 10534-2. Acoustics – Determination of sound absorption coefficient and impedance in impedance tubes. Part 2: Transfer-function method. 1st edition. 1998.

[7] Munjal M I Acoustics of Ducts and Mufflers. John Wiley & Sons. 2nd edition, 2014.

A COMPUTATIONAL STUDY OF DEPTH OF VIBRATION INTO VOCAL FOLD TISSUES

Anil Palaparthi, MS[1], Simeon Smith, MS[1], Ted Mau, MD PhD[3], Ingo R. Titze, PhD[1,2]

[1] National Center for Voice and Speech, University of Utah, Salt Lake City, UT USA
[2] Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, IA, USA
[3] University of Texas Southwestern Medical Center, Dallas, Texas, USA

***Keywords*: Vocal Folds; Effective Depth; Fiber-Gel Finite Element Model**

Abstract:

Objectives: Effective depth of vocal fold vibration is the depth of vocal fold tissue that is in motion during vocal fold vibration. The goal of this study is to quantify this depth systematically under various phonatory conditions using a finite element fiber-gel vocal fold model.

Methods: The parameters varied to simulate different phonatory conditions were lung pressure, thyroarytenoid activation, and cricothyroid activation. The horizontal and vertical excursions of each finite element nodal point trajectory were recorded to compute their trajectory areas. Extent of vibration was then studied based on the variation of trajectory areas as a function of depth in several coronal sections along the anterior-posterior direction. These results were compared with the medial surface results from excised human hemi-larynx experiments published in Dollinger et al. (2006, 2016).

Results: The results suggested that the vocal fold nodal trajectory area decreases quadratically as a function of depth for lower CT activation values. When the CT activation is high, the vocal fold nodal trajectory area has a complex behavior as a function of depth due to the presence of a stiff ligament. The effective depth was then quantified as the depth at which the area is half of the summed trajectory area.

Conclusions: The vocal fold trajectories on the medial surface obtained from Fiber-Gel finite element model were comparable to those observed in excised human larynx experiments. The effective depth decreased with thickness in the superior-inferior direction, but remained constant in the anterior-posterior direction.

RELIABILITY OF RELATIVE FUNDAMENTAL FREQUENCY AND CONVENTIONAL ACOUSTIC AND AERODYNAMIC MEASURES IN INDIVIDUALS WITH HEALTHY VOICES

Yeonggwang Park[1], Cara E. Stepp[1,2,3]

[1] Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA
[2] Department of Biomedical Engineering, Boston University, Boston, MA, USA
[3] Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, MA, USA

Keywords: Reliability of Voice Assessment; Relative Fundamental Frequency; Acoustic and Aerodynamic Measures

Introduction

Recent studies have shown that excessive laryngeal tension and vocal effort associated with vocal hyperfunction, Parkinson's disease, and other functional and neurological voice disorders can be assessed with an acoustic measure, Relative Fundamental Frequency (RFF). However, the reliability of RFF has not been evaluated. This study presents an analysis of the reliability of RFF in individuals with healthy voices and a comparison of reliability between RFF and conventional measures of voice.

Methods

Acoustic and aerodynamic measurements were performed on 28 individuals with healthy voices on five consecutive days. Participants produced RFF stimuli, a sustained /ɑ/, and a reading passage to allow for extraction of acoustic measures, and /pa/ trains to allow for extraction of aerodynamic measures. Additional information such as wake-up time, voice discomfort level, and the inability to produce soft voice (IPSV) was recorded to detect possible voice changes. The reliabilities of RFF and conventional measures across the multiple recordings were assessed with intraclass correlation coefficients (ICCs).

Results and Discussion

Moderate reliabilities (offset cycle 10: ICC=0.71; onset cycle 1: ICC=0.64) were found for RFF values. The estimated ICCs for RFF were lower than for mean vocal fundamental frequency (ICC=0.99), smoothed cepstral peak prominence (CPPS; vowel: ICC=0.87; sentence: ICC=0.76), shimmer (ICC=0.91), harmonics-to-noise ratio (HNR; ICC=0.91), and mean airflow rate (comfortable: ICC=0.78; loud: ICC=0.84). The ICCs for RFF were similar to the ICCs for jitter (ICC=0.67) and phonation threshold pressure (PTP; ICC=0.74) and higher than the ICCs for subglottal pressure estimates (comfortable: ICC=0.51; loud: ICC=0.61). Most of the conventional measures examined are intended to measure perceived overall dysphonia, usually caused by vocal changes related to the structure of the vocal folds, whereas RFF is thought to reflect perceived strain related to laryngeal tension, a functional aspect of vocal production. This could explain why RFF showed only moderate reliability. The ICC for RFF onset cycle 1 measured with soft voice were higher (ICC=0.76) than for RFF onset cycle 1 measured using comfortable and loud voices (comfortable: ICC =0.64; loud: ICC=0.50). This suggests that vocal effort level may have been more consistent in soft voice trials than during use of comfortable or loud voice.

Conclusion

This investigation has revealed that RFF has comparable reliability to conventional measures of voice. This expands the potential for clinical application of RFF as a reliable assessment of laryngeal tension.

Acknowledgments

IMMEDIATE EFFECT OF ELECTROSTIMULATION IN SEMIOCCLUDED VOCAL TRACT EXERCISES (SOVTs) IN CHILLÁN TEACHERS.

Jazmín Pérez-Serey[1], Alejandro Núñez[2], Francisca Carrasco[3], Sebastián Dinamarca[3], Carolina Postel[3], Yenny Soto[3], Bárbara Toro[3]

[1] Academic Director Speech therapy, Universidad Pedro de Valdivia, Chile; Postgrado, Universidad Adventista de Chile.
[2] Teacher Speech Therapy, Universidad Pedro de Valdivia, Chillán, Chile.
[3] Speech Therapy, Universidad Pedro de Valdivia, Chillán, Chile.

Keywords: **Acoustic parameters, electro-stimulation, semi-occluded vocal tract, teachers, voice**

INTRODUCTION

There are several therapies whose effectiveness has been proven to preserve teacher's voice by semi-occluded vocal tract technique (SOCTs) [3], which has benefits such as to increase in source-filter interaction, to obtain more efficient and economical voice, to promote high pressures in the vocal tract, to allow an amplification of the sensation of internal vibration, to increase the perception of less resistance in the passage of sound through the vocal tract and to eliminate record breaks [3,4,5-9]. According to currently emerging therapies, there is a Electro Stimulation (ES) therapy, where the main effects are activation, analgesia, relaxation of the circulatory, muscular and sensory systems. It also contributes to muscle toning [1,6]. However, both therapies in a complementary way have not been studied. So the question arises is: Does the application of electro-stimulation in semi-occluded vocal tract exercises produce immediate improvements in teachers' acoustic parameters? In this way, the objective of this study is to determine the immediate effect of electrostimulation in semi-occluded vocal tractor exercises on acoustic parameters in teachers from Chillán.

METHODS

Human Data

A descriptive-comparative study of quantitative type and quasi-experimental design was conducted with 68 teachers who are in the range of 28 to 60 years old. There was a randomly selection for control and experimental samples. Finally 30 individuals decided to voluntarily participate in the study and achieved the inclusion criteria and signed the informed consent.

The criteria for inclusion include having a teaching degree from a Chilean university, having at least 2 years of work experience in teaching, having at least 30 teaching hours per week and signing informed consent.

The criteria for exclusion include being in the age range below 28 years and over 60 years, teaching over 30 years, having less than 30 hours of weekly workload, presenting basic pathology, dysfunction and / or diagnosed organic, presenting grave or chronic respiratory pathology, having formal vocal and / or vocal therapy studies, having abnormal acoustic parameters in initial evaluation, presenting severe cardiac alterations, being a pacemaker carrier, presenting cutaneous hypersensitivity and being pregnant.

Instrumentation and Measurement

An evaluation of the voice in a control and experimental group was carried out with the Acoustic Analysis Program PRAAT and Maximum Time of Phonation (TMF) before and after vocal therapy at the beginning of the working day. The semi-occluded vocal tract therapy with light bulb was applied to the control group. The exercises lasted 2.5 minutes according to the following sequence [3]:

1) Tube phonation producing sustained and comfortable frequency for each subject in intensity and height.

2) Tube phonation producing ascending and descending glissandos.

3) Tube phonation making accents of intensity and frequency with use of the abdominal muscles. The realization of the accents was ascending.

4) Production of melody chosen by the user inside the tube.

For the experimental group, semi-occluded vocal tract therapy was applied in conjunction with electrostimulation therapy, which consisted of applying current between two electrodes through the tissue. The current was supplied in a two-phase pulse in one direction, followed by a brief pause and then returning in the opposite direction. The electro-stimulator was programmed at a preset frequency of 30 Hz. The electrodes were placed horizontally in the cricothyroid space. The intensity was increased until having a feeling of tightening using a maximum intensity of 10 milliamps. The intensity is maintained at that level to begin vocal training. [1].

Analysis

Descriptive statistics were applied through mean and standard deviation. For demonstrating normality was used Shapiro Wilk, for homogeneity the Levene test was applied and for approving and rejecting hypotheses T-Student was used for two independent samples.

RESULTS

Table N°1 shows that all parameters of the control group decreased the normality ranges. Jitter is <1.040%, Shimmer <3.810, HNR> 20dB, TMF> 15 seconds.

Table N° 1: Descrptive analysis of Control Group before and after interventión.

	Media Pre	Desv. Stand. Pre	Media Post	Desv. Stand. Post	Valor p
JITTER	0,88	0,08	0,79	0,08	*0,000
SHIMMER	2,93	0,38	2,75	0,35	*0,014
HNR	21,44	1,11	22,27	1,16	*0,012
TMF	16,93	1,39	18,60	1,45	*0,000

* p<0,05

In Table Nº 2 it is observed that the parameters of the experimental group decreased more widely than in the experimental group.

Tabla N° 2: Descrptive analysis of Experimental Group before and after interventión.

	Media Pre	Desv. Stand. Pre	Media Post	Desv. Stand. Post	Valor p
JITTER	0,85	0,08	0,69	0,05	*0,000
SHIMMER	2,82	0,40	2,46	0,23	*0,015
HNR	21,18	0,87	23,27	8,86	*0,013
TMF	16,87	2,12	22,20	2,37	*0,000

* p<0,05

Table Nº 3 shows significant differences in all vocal parameters between control and experimental groups; the experimental group means stand out above the control.

Tabla N° 3: Comparative Analysis Post Intervention Student's T test for independent samples.

Parámetros Acústicos	Grupo Control Post intervención	Grupo Exp. Post intervención	P Valor
	Media	Media	
JITTER	0,79	0,69	*0,002
SHIMMER	2,75	2,46	*0,048
HNR	22,27	23,27	*0,018
TMF	18,60	22,20	*0,000

* p<0,05

DISCUSSION

The existing differences in pre and post intervention in semi-occluded vocal tract therapy with electrostimulation in teachers, agree with Guimarães's research (2010) [1], which indicates that post-application of electrostimulation frequency disturbance (Jitter) decreased; in the disturbance of the amplitude wave (Shimmer) an improvement in the relative value occurred, and the result of the HNR parameter increased in all cases, being able to indicate greater relaxation at the laryngeal level. In contrast,

Gorham, Fowler and Hapner (2010) indicate that after the application of transcutaneous electrostimulation (TENS) to 30 individuals without vocal pathologies between 19 and 59 years the values of the acoustic parameters did not show significant differences after intervention with electrostimulation, with the exception of the Shimmer parameter, which decreased by 7%.[2] According to Oliveira, Cortés, Alves and Cordeiro (2015), there are beneficial results regarding the use of electrostimulation in individuals without vocal alteration.[5]

CONCLUSION

Comparing the acoustic parameters in both groups of pre and post intervention study, it was found that in both groups there were variations in the evaluated parameters, but the greatest significance was in the group that received intervention of both therapies in parallel.

REFERENCES

[1]Guimarães BT. Furkin AM, Silva RG. Eletroestimulação neuromuscular na reabilitação dea disfagia orofaríngea. *Rev.so. bras. Fonoaudiol.,* 2010;15(4):615-21.

[2]Gorham-Rowan M, Fowler L yHapner E. acoustic Analysis of Voice change in normal Speakers Following Transcutaneous Electrical Stimulation to the Laryngeal Area. *The Reabilitation Journal,* 2010;3(1):67-74.

[3]Guzman M. Voice Therapy with semi-occuded voccal tract: a case study. *Revista chilena de Fonoaudiología,* 2012;11:87-97.

[4]Guzman M, Laukkanen AM, Krupa P,Horácëk J, Svec J. Vocal Tract and Glottal Function During and After Vocal Exercising With Resonance Tube and Straw. Journal of Voice, 2013; 27(4):524-534.

[5]Oliveira S, cortés A, Alves K, Cordeiro N. Uso da electroestimulação na ckínica fonoaudiológica: uma revição integrativa da literatura. Revista CEFAC, 2015;17(5):1620-1632.

[6]Santos JK, Gama AC, Silvério KC, Oliveira NF. The use of electrical stimulation in speech therapy clinical: an integrative literatura review. *Rev CEFAC* 2015; 17(5):1620-1632.

[7]Sampaio M. Oliveira, G. Behlau, M. Investigation of immediate effects of two semi-ocluded vocal tract exercises. *Pro-Fono revista de actualización científica,* 2008;20, 261-266.

[8] Titze & Laukkannen AM. Can vocal econocmy en phonation be increased with an artificially lengthened vocal tract? A computer modeling study. Logopedics Phoniatrcs Vocology. 2007;32:147-156.

[9] Titze, I. Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings. *J Speech Lang Hear Res* 2006; 49, 448–459.

DIFFERENCES IN VOICE PARAMETERS IN THOSE WHO EXPERIENCE A BIOLOGICAL STRESS RESPONSE AND THOSE WHO DO NOT

Brittany L. Perrine, Ph.D.[1,2], Ronald C. Scherer, Ph.D.[1]

[1] Department of Communication Sciences and Disorders, Bowling Green State University, Bowling Green, Ohio, USA
[2] Department of Communication Sciences and Disorders, Baylor University, Waco, Texas, USA

***Keywords*: e.g. Voice; Psychosocial Stress; Aerodynamics; Acoustics**

INTRODUCTION

Psychological stress is implicated in the development and progression of a number of diseases and has been suggested to play a part in the development of voice disorders [e.g., 1]. Two main stress systems of the body control the biological reaction to stress: the sympathetic nervous system and the hypothalamic-pituitary-adrenal (HPA) axis. Activation of the former stress system has resulted in no significant vocal acoustic changes [2-3] while studies of the influence of HPA axis activation on the voice have noted an increase in fundamental frequency [4-6] and a decrease in fundamental frequency [7]. However, none of the studies examining HPA axis activation have measured a marker of the HPA axis (i.e., cortisol) in the participants to ensure that the stress system was activated.

The aim of this study was to determine if there are differences in vocal aerodynamic and acoustic production parameters between participants who experience HPA axis activation and those who do not.

METHODS

Participants

Nineteen female participants with an age range of 18 to 23 years old (mean: 18.89; standard deviation: 1.45; median: 18) were included in this study. No participant was on a hormonal birth control method. Those who had indications in their health history of an artificial increase or decrease in salivary cortisol (e.g., gingivitis, recent pregnancy, smoking or nicotine use, oral or topical steroid use, asthma, allergies, etc.) were excluded from participating.

Instrumentation and Measurement

Aerodynamic recordings were made using the Glottal Enterprises aerodynamic system (MSIF-2 S/N 2049S) and acoustic recordings were made using a condenser microphone attached to a headset placed 4 cm and 45 degrees to the left side of the mouth (C 420 III PP MicroMic from AKG Acoustics; frequency response 20-20,000). The signals were simultaneously digitized at 20,000 samples per second for each channel to a 16 bit DATAQ A/D converter system (Model DI-2108 Series).

Saliva samples were collected using the SalivaBio Oral Swab system (Salimetrics, State College, PA). The swab was placed under the tongue of the participant for 2 minutes to collect approximately 1 to 2 mL of saliva. Saliva samples were immediately labeled and placed in a -20° C freezer for storage until analysis (maximum time: 4 months).

Protocol

All recordings were made from approximately (3 to 5 pm) to control for the circadian rhythm of cortisol. The Trier Social Stress Test (TSST) [8] was administered following the protocol presented in [9]. Briefly, a saliva sample and voice recording (which included five to seven sets of repetitions of /pa/ seven to nine times on one breath and the Rainbow Passage) occurred in series (1) when the participant arrived, (2) after a 10 minute quiet sitting period, (3) after 10 minutes of preparing for a public speech, (4) immediately after a 10 minute public speaking and mental arithmetic task in front of an audience, (5-7) and every 10 minutes thereafter while the participant quietly sat in a room between recordings, for a total of 7 measurement time points.

Analysis

Saliva samples were analyzed by the University of Michigan Core Assay Facility using the Salimetrics Cortisol ELISA Kit to measure salivary cortisol in the sample (inter-assay coefficient: 6.13% at 2.700843 nmol/l to 2.68% at 27.07469 nmol/l; intra-assay coefficient of variance: 5.48%). Two samples did not have enough saliva for a salivary cortisol analysis. If a 2.5 nmol/l increase in salivary cortisol was noted after the stressor, the HPA axis was considered activated [10].

The mean and standard deviation of the fundamental frequency (f_o) were taken from the second sentence of the Rainbow Passage using the pulse analysis feature in Praat Software (Version 6.0.14). Pulses were removed from the entire syllable if vocal fry was noted anywhere in the syllable to ensure the mean fundamental frequency represented only modal register. The average oral airflow and estimated subglottal pressure (Psub) were extracted from the /pa/ syllables using custom Matlab software (Sigplot).

A two-way repeated measures ANOVA was conducted for each dependent variable and effects were interpreted using a significance of $p < .05$.

RESULTS

Ten of the 19 participants had an HPA axis response (a 2.5 nmol/l increase in salivary cortisol from before the stressor). There were no significant interactions between HPA axis response group and measurement time for average airflow or estimated Psub. There was a significant interaction between HPA axis response group and measurement time on speaking f_o, which held for the HPA axis response group in follow-up analysis. Pairwise comparisons indicated that speaking f_o was significantly higher at time 3 than time 5 by 0.57 semitones and was higher at time 4 than time 6 by 0.52 semitones for the HPA response group (Fig 1). There was a surprisingly significant main effect of estimated Psub between groups, with the no HPA axis response group presenting with a higher estimated Psub (c. 1.6 cm H2O) at all times (Fig 2).
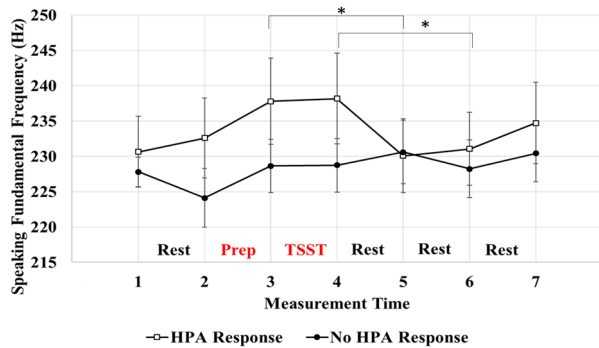

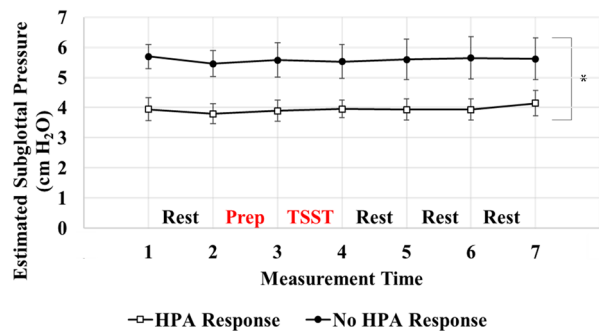**Fig. 1: Mean speaking fundamental frequency over the experimental tasks for each group.**

Fig. 2: Estimated subglottal pressure over the experimental tasks for each group.

DISCUSSION AND CONCLUSION

There were no differences in average airflow and estimated subglottal pressure in those who had an HPA axis response and those who did not from before to after the stressor (over time). As expected from past literature, the speaking f_o was slightly higher around the time of the stressor; however, this increase was only seen in those who had an HPA axis response. A higher estimated subglottal pressure was seen at all times (before and after the stressor; not related to the stressor) in the group that did not have an

HPA axis response. This was associated with a higher laryngeal airflow resistance. The participants who did not have an HPA axis response to the stressor may have used behavioral adaptations (increasing the lung pressure) to avoid voice quality changes that are more apparent at lower lung pressures. These results suggest that there may be underlying differences in voice production parameters overall, not just in response to the stressor, in those who have an HPA axis response to stress and those who do not.

REFERENCES

[1] Aronson, AE. *Clinical Voice Disorders: An Interdisciplinary Approach.* 3rd ed. New York, NY: Thieme; 1990.

[2] Giddens, CL, Barron, KW, Clark, KF, Warde, WD. Beta-adrenergic blockade and voice: A double-blind placebo-controlled trial. *J Voice.* 2010;24(4):477-489.

[3] Alvear, RMD, Barón-López, FJ, Alguacil, MD, Dawid-Milner, MS. Interactions between voice fundamental frequency and cardiovascular parameters: Preliminary results and physiological mechanisms. *Log Phon Voc.* 2013;38(2):52-58.

[4] Johannes, B, Wittels, P, Enne, R, Eisinger, G, Castro, CA, Thomas, JL, … Gerzer, R. Non-linear function model of voice pitch dependency on physical and mental load. *Euro J App Phys.* 2007;101(3):267-276.

[5] Rothkrantz, LJ, Wiggers, P, van Wees, JWA, van Vark, RJ. Voice Stress Analysis. In *Int Conf Text, Speech, and dialogue*. 2004;449-456.

[6] Tse, ACY, Wong AWK, Whitehill, TL, Ma EPM, Masters, RS. Analogy Instruction and Speech Performance under Psychological Stress. *J Voice*. 2014;28(2):196-202.

[7] Dietrich M. *The effects of stress reactivity on extralaryngeal muscle tension in vocally normal participants as a function of personality*. (Doc Diss, U of Pittsburgh).

[8] Kirschbaum, C, Pirke, KM, Hellhammer, DH. The 'Trier Social Stress Test'- A Toll for Investigating Psychobiological Stress Response in a Laboratory setting. *Neuropsychobiology*. 1993;28(1-2):76-81.

[9] Perrine, B. *The Influence of Stress on the Voice*. (Doc Diss. BGSU).

[10] Weitzman, ED, Fukushima, D, Nogeire, C, Roffwarg, H, Gallagher, TF, & Hellman, L. Twenty-four Hour Pattern of the Episodic Secretion of Cortisol in Normal Subjects. *J Clin Endo & Metab*. 1971;33(1):14-22.

A FINITE ELEMENT STUDY OF A REALISTIC VOCAL FOLD POSTURING MODEL AND ITS COUPLING WITH 1D FLOW STRUCTURE INTERACTION

Ngoc Pham[1], Biao Geng[1], Qian Xue[1], and Xudong Zheng[1]

[1] Department of Mechanical Engineering, University of Maine, Orono, Maine, USA

***Keywords*: 3D finite element, active muscle contraction, vocal fold posturing**

INTRODUCTION

Quantification of passive and active properties of laryngeal muscles is critical for vocal fold simulation, as they govern vocal fold adduction/abduction. This work aims at using a 3D finite element method to simulate the passive and active response of a canine vocal folds, and to study the shape of the vocal folds under the contraction of different vocalis muscles or groups of muscles. A model for fiber-reinforced materials was employed to simulate stretching-induced stress, while a Hill-based contractile element was used to capture activation-induced stress. A three-dimensional muscles and cartilages model was also developed for complete vocal fold posturing simulation.

METHODS

The cartilages (i.e., cricoid, thyroid, and arytenoid) used for simulations in this work were reconstructed from a MRI scan, taken from a canine. Also reconstructed from the MRI scan were thyroarytenoid (TA), lateral cricorytenoid (LCA), and posterior cricoarytenoid (PCA) muscles. Cricothyroid and interarytenoid muscle were approximated using simplified geometries, since their real anatomy was unable to obtain from the MRI scan. Contraction of these muscles are critical in vocal fold posturing, as TA/IA/LCA muscles are vocal adductor, PCA muscles are vocal fold abductor, and CT muscles create tension on the vocal fold.

Passive properties of the laryngeal muscles were modeled using a model for fiber-reinforced materials, given as follows:

$$U = C_{10}(\bar{I}_1 - 3) + \frac{1}{D_1}(J-1)^2 + \frac{k_1}{2k_2}\left[e^{k_2(\bar{I}_4-1)^2} - 1\right] \quad (1)$$

where $\bar{I}_1$ and $\bar{I}_4$ is the first and the fourth invariant of the reduced Cauchy-Green tensor, respectively, J is the Jacobian determinant of the deformation, C_{10}, k_1, and k_2 are the material constants, and D_1 is the compressibility factor. The material constants for each muscle (see Table 1) were obtained through a parameter sensitivity analysis, which ensured reasonable matching between its experimental stress-stretch curves with numerically-generated stress-stretch curves in both static and dynamic conditions. To describe muscle active properties, a Hill-based contractile model was used. It is formulated as follows:

$$\sigma_A = a(t)\sigma_{max}\sigma_{TL}^*\sigma_{TV}^* \quad (2)$$

where σ_A is the active stress, a(t) is the activation level, σ_{max} is the maximum active muscle stress, and σ_{TL}^* and σ_{TV}^* are two scaling factors, which scale the active stress with muscle stretch and stretch rate, respectively [1]. Equation (1) and (2) were coupled in a finite element package (i.e., CalculiX) to simulate concurrent passive and active response of the muscles.

RESULTS AND DISCUSSION

Table 1 presents three different sets of the material constants, used in Eq. (1), which describe passive properties of the muscles of interest reasonably.

Table 1: Material constants for passive properties

Muscle	C_{10} [kPa]	k_1 [kPa]	k_2 [-]
CT&TA	3	0.15	9
LCA&PCA	2.7	2	5.5
IA	2.035	10	2.5

It is apparent from Table 1 that CT and TA muscles share the same set of material constants, since their passive response to stretching are almost identical [2]. Similarly, passive properties of LCA and PCA muscles can be described by a common set of material constants because of the same reason.

To validate the coupling between Eq. (1) and Eq. (2), the force obtained from a 4-cycle stretch-release experiment by Alipour-Haghighi et al. [3] was used. A rectangular bar, measure $15 \times 3 \times 3$ mm^3 was employed in the simulation to represent the muscle sample. In the simulation, the boundary conditions, the number of stretching cycles, the stretch rate, and the activation time were identical to those of the experiment for force comparison. Such comparison is shown in Figure 1.
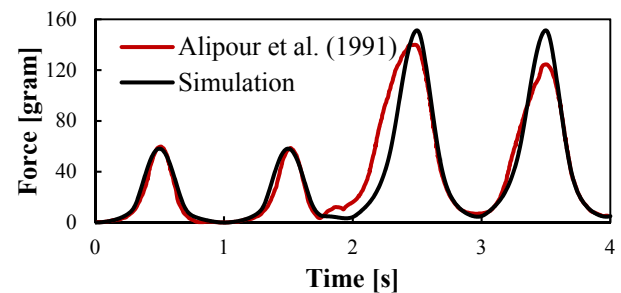

Figure 1: Agreement of cyclic force, computed from the coupling with that from experimental measurement.

In general, the experimental data can be reproduced by coupling Eqs. (1) & (2), despite the noticeable difference in the last two peaks, where muscle activation was enabled. In all four cycles, the shape, the height, and the position of the force peaks predicted by the coupling model are comparable to the experimental measurement.

Shown in Figure 2 is the 3D muscles and cartilages model developed in this work. In the 3D model, the cartilages and the laryngeal muscles were reconstructed from the MRI scan, except CT and IA muscles, whose geometries were simplified.
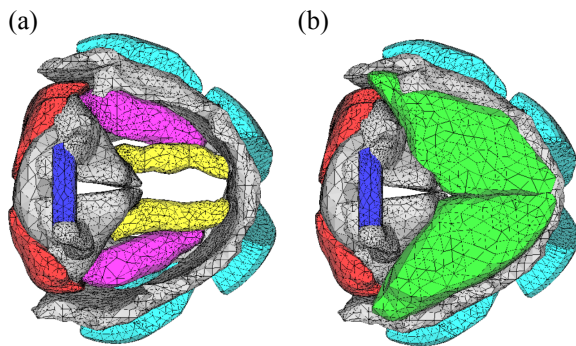

Figure 2: Three-dimensional muscles and cartilages model. (a): superior view without fat layers; (b): superior view with fat layers. The color code is as follows: cartilages = gray, CT muscles = turquoise, TA muscles = yellow, IA muscles = blue, LCA muscles = magenta, PCA muscles = red, and fat layers = green.

Although the representation of CT and IA muscles are simplified geometries, the orientation, length, and mass of the muscle match their anatomic characterization reported in literature [4; 5]. Results from the MRI scan suggest that, besides the TA and LCA muscles, other components (e.g., other muscle bundles of same type or different muscle types) should present in the space between the arytenoid and the thyroid cartilage. However, due to low resolution of the MRI scan, those components remain undetermined. Hence, under our consideration, the materials in that space disregarding the TA and LCA muscles are fat layers (colored in green in Fig. 2b). Complete reconstruction of a 3D muscles and cartilages model enables the capability of simulating different combinations of muscle activation (individual, in a pair, or in a group). One of which is shown in Figure 3.

Figure 3 presents the vocal fold strain when CT and TA muscles are activated at the same time at different activation levels. It clearly indicates that the vocal folds are compressed when only TA muscles are activated at maximum level (see Figure 4 for the model deformation), while the vocal folds are stretched if only CT muscles contract at maximum level. Concurrent activation of TA and CT muscles at different activation levels results in highly nonlinear deformation of the vocal folds.
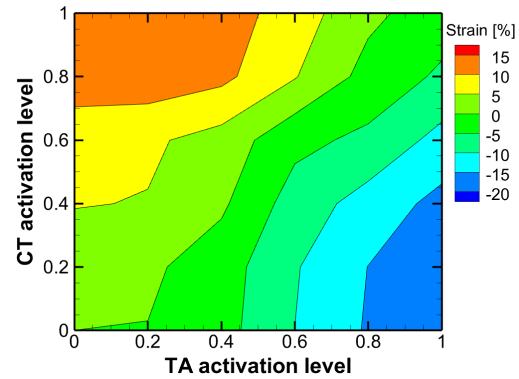

Figure 3: Contour of vocal fold strain at different activation levels of TA and CT muscles.
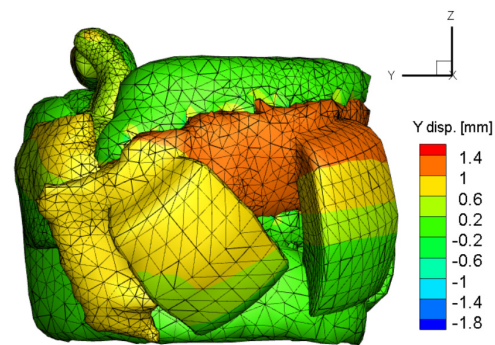

Figure 4: Model deformation under maximum activation of TA muscles and zero activation of CT muscles. The color indicates the displacement in Y direction.

CONCLUSION

A Hill-based contractile model is successfully coupled with a fiber-reinforced model in a three-dimensional finite element study to simulate the passive and active response of a canine vocal folds. Different combinations of muscle activation lead to different vocal fold deformation levels.

ACKNOWLEDGMENTS

REFERENCES

[1] Smith, S.L., and Hunter, E.J. (**2014**), J. Acous. Soc. Am. **135**, 2041-2051.

[2] Hunter, E.J., and Titze, I.R. (**2007**), J. Appl. Physiol. **103**, 206-219.

[3] Alipour-Haghighi, F., Perlman, A.L., and Titze, I.R. (**1991**), Ann. Otol. Rhinol. Laryngol. **100**, 626-631.

[4] Cox, K.A., Alipour, F., and Titze, I.R. (**1999**), Ann. Otol. Rhinol. Laryngol. **108**, 1151-1158.

[5] Mineck, C.W., Chan, R., Tayama, N., and Titze, I.R. (**2000**), Ann. Otol. Rhinol. Laryngol. **109**, 505-513.

VOCAL FOLD VIBRATORY OUTCOMES FOLLOWING HIGH DOSE OF GLUCOCORTICOIDS

Maria Powell, PhD, CCC-SLP[1]; Lea Sayce, DPhil[1]; Emily E. Kimball, MS[2]; Shintaro Sueyoshi, MD[1]; Gary Gartling, MS[2]; Bernard Rousseau, PhD, MMHC, CCC-SLP[1-3]

[1] Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
[2] Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, Tennessee, USA.
[3] Department of Mechanical Engineering, Vanderbilt University, Nashville, Tennessee, USA.

Keywords: Glucocorticoid Steroids; Vocal Folds; Vibration; High-Speed Videoendoscopy

INTRODUCTION

Administration of glucocorticoid steroids (GCs) is an increasingly common treatment for a number of benign laryngeal disorders [1]. Despite the ubiquitous use of GCs for treating these voice disorders, evidence supporting their safety and efficacy is lacking. Positive treatment effects following a course of GCs have been reported, including reduction in lesion size [2], resolution of vocal nodules [3] or prevention of surgical intervention for benign phonotraumatic lesions [4]. However, these retrospective case-series apply variable methodologies and lack sufficient controls, bringing the validity and/or generalizability of findings into question. Further, adverse side effects, including muscle atrophy [4-6], plaque deposits [5,6], and mucosal changes have also been reported [7]. These inconsistencies result in a lack of consensus on the role of GCs in the treatment of voice disorders. In the absence of clear, empirical evidence for informing drug selection, dosing, and route of administration for treating laryngeal disorders, otolaryngologists report their decision-making framework is primarily based on past experience and drug familiarity [1]. In order to support the ongoing use of GCs to treat benign voice disorders, the molecular, structural, and functional response to this treatment must be considered. As part of a larger study examining the safety and efficacy of GC use for the treatment of phonotrauma, the purpose of this study is to systematically investigate the effect of a 6-day high-dose course of two commonly prescribed GCs (dexamethasone and methylprednisolone) on vocal fold vibratory outcomes in an *in vivo* rabbit phonation model.

METHODS

All procedures were approved by the Vanderbilt University Medical Center Institutional Animal Care and Use Committee and performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animal Welfare Act (7 U.S.C. et seq.)

Glucocorticoid Steroid Treatment

Fifteen male, age and weight matched, New Zealand white breeder rabbits were randomized to receive a 6-day course of either 450µg dexamethasone (n=5), 4.5mg methylprednisolone (n=5), or saline (n=5). Injections were volume matched at 0.225ml and administered into alternating hind quarters. Following a 14 day recovery period, all animals underwent an *in vivo* phonation procedure.

In Vivo Phonation Procedure

Specifics of this procedure are detailed elsewhere [8]. Briefly, the animal was anesthetized with a continuous infusion of ketamine and dexmedetomidine. The animal was shaved, and dissected to the level of the trachea/larynx from the submentum to sternal notch. The trachea was transected, and oxygen was administered to the lungs via an uncuffed endotracheal tube. The upper airway was sealed by a cuffed endotracheal tube, and continuous humidified airflow was delivered though the glottis. Vocal fold adduction to produce phonation was achieved by electrical stimulation of stainless-steel hooked electrodes inserted bilaterally into the cricothyroid membrane and cricothyroid muscles.

The animal was suspended using a side-slotted laryngoscope, and the larynx was visualized using a 0° 4.0-mm rigid endoscope coupled to a FASTCAM MC 2.1 monochrome high-speed camera and 300W continuous Xenon light source. Animals were briefly phonated at modal intensity followed by raised intensity. Vocal fold vibration for each phonation type was recorded at 8000 frames per second (fps) with a spatial resolution of 128x512 pixels. Following phonation, animals were euthanized by pentobarbital overdose.

High-Speed Videoendoscopy Analysis

Modal and raised intensity high-speed image sequences were visually inspected, and a representative 200-frame (25ms) video segment was selected for analysis. These 200-frame videos were upsampled to 24000 fps using cubic spline interpolation and imported into a custom Matlab software for digital kymography extraction and edge detection using the Canny edge detector algorithm (Fig. 1) [8]. Kymograms were quantitatively analyzed for amplitude, amplitude perturbation (cycle-to-cycle amplitude variation), and lateral phase difference.
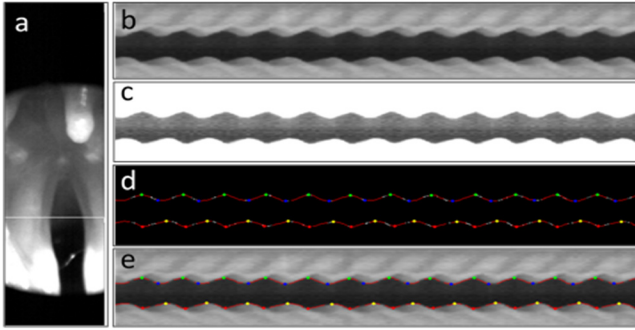
Fig. 1: Image analysis of high-speed videoendoscopy data. *Using the full image sequence as reference (a), the point of maximum medio-lateral tissue deformation (amplitude) was identified, and corresponding digital kymographic line scan was displayed (b). The intensity of the image was manually adjusted to maximize contrast between the tissue and glottis (c). Segmentation was conducted on this high-contrast image (d), and the resulting edge detection was compared to the original line scan for accuracy (e).*

Fig.2: *Mann-Whitney U post-hoc comparisons show increased amplitude for dexamethasone and methylprednisolone-dosed animals relative to saline control. Methylprednisolone-dosed animals also had increased amplitude at raised intensities compared to modal. Findings are significant at p<0.05.*

RESULTS

A two-tail Wilcoxon test showed a significant main effect of GC on amplitude (p=0.0379). Post hoc Mann-Whitney U tests revealed both dexamethasone (p=0.0404) and methylprednisolone-dosed (p=0.0189) animals had approximately 6.5% greater amplitude at raised intensity compared to saline controls (Fig. 2). Methylprednisolone-dosed animals also had greater amplitude at raised versus modal intensities (p=0.0234). There was a main effect of intensity on amplitude perturbation (p=0.0268); however, post hoc analysis revealed these increased perturbations at raised intensity were not treatment-specific. Lateral phase difference was not significant for treatment group or phonation condition.

DISCUSSION AND CONCLUSIONS

Long term changes in vocal fold vibration were noted in both dexamethasone and methylprednisolone-treated animals. Amplitude was subtly increased in the raised intensity condition for both GC-treated groups compared to saline controls. This phenomenon may be more pronounced in the methylprednisolone-treated animals, as significant differences in amplitude between modal and raised intensity phonation was present within this group as well. The mechanism behind this altered vibratory function is unclear; however, it is possible that the increased amplitude could be secondary to muscle atrophy, which has been reported in clinical literature following GC treatment [4-6]. Importantly, in the current study, vibratory function was investigated fourteen days following the final GC injection. Future studies should investigate changes in vibratory function immediately following the same course of treatment.
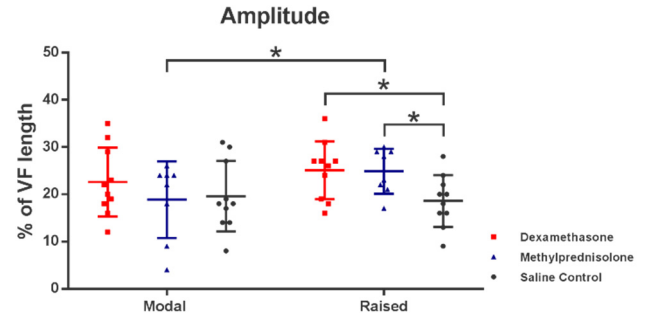
REFERENCES

[1] Govil N, Rafii BY, Paul BC, Ruiz R, Amin MR, Branski RC. Glucocorticoids for vocal fold disease: A survey of otolaryngologists. *J Voice*. 2014;28(1):83-87.

[2] Wang C-T, Liao L-J, Lai M-S, Cheng P-W. Comparison of benign lesion regression following vocal fold steroid injection and vocal hygiene education. *Laryngoscope*. 2014;124(2):510-515.

[3] Tateya I, Omori K, Kojima H, Hirano S, Kaneko K, Ito J. Steroid injection to vocal nodules using fiberoptic laryngeal surgery under topical anesthesia. *Eur Arch Oto-Rhino-Laryngol*. 2004;261(9):489-492.

[4] Mortensen M, Woo P. Office steroid injections of the larynx. *Laryngoscope*. 2006;116(10):1735-1739.

[5] Lee S-H, Yeo J-O, Choi J-I, et al. Local steroid injection via the cricothyroid membrane in patients with a vocal nodule. *Arch Otolaryngol Head Neck Surg*. 2011;137(10):1011-1016

[6] Lavy JA, Wood G, Rubin JS, Harries M. Dysphonia associated with inhaled steroids. *J Voice*. 2000;14(4):581-588.

[7] Andrade Filho PA, Rosen CA. Vocal fold plaque following triamcinolone injection. *Ear Nose Throat J*. 2003;82(12):908, 911.

[8] Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell*. 1986;8(6):679-698.

MEASURES OF VOICE ONSET TIME: A METHODOLOGICAL STUDY

Rebecca Rae[1], Ronald Scherer[1]

[1] Communication Sciences and Disorders, Bowling Green State University, Bowling Green, OH, U.S.A

***Keywords*: e.g. Voice Onset Time; Methodology; Aerodynamics; Electroglottography**

INTRODUCTION

There have been numerous studies of voice onset time (VOT) through spectrographic, oscillographic (microphone), and glottographic measurement approaches. There appears to be only one previous research study that has taken into account airflow as a signal to accurately measure the burst and phonatory onset for VOT measurement [1]. It is possible that the most accurate measure of when the vocal folds begin their oscillation during phonatory onset is revealed through the airflow signal, because airflow may be more sensitive to this initial oscillation by showing the modulation of the airflow than the acoustic signal is in showing acoustic modulation.

The current study comparatively examined four different measurement approaches (using the audio signal, spectrographic display, electroglottographic (EGG) signal, and averaged wideband airflow) for the measurement of voice onset time. This study introduces the phrase "initial detection of phonation, IDP", which is the moment in time for the initial indication of vocal fold or glottal oscillation in the waveform of the acoustic, airflow, or electroglottographic signals. This is visualized as the initial modulation of the signal when it appears to have a strongly related period of oscillation related to the immediately subsequent voicing cycles.

METHODS

Participants

Four healthy participants (2 adult males and 2 adult females) were recruited for the current study. All participants were native speakers of English, had normal voice and speech, reported no history of receiving training in professional speech, voice, or singing, and reported no history of smoking or current illness.

Speech Stimuli

Each of the six English stop consonants (/p, b, t, d, k, g/) were paired with the vowel /ɑ/ and produced in the carrier phrase "CV-the-CV" (e.g. "pɑ the pɑ") spoken with equal stress on each of the CV syllables. This phrase was selected in order to attempt to equate the stress level of the first syllable across the utterances.

Instrumentation and Analysis

The wideband airflow, microphone, and EGG signals were simultaneously recorded using a DATAQ A/D converter (DI-2108 Series, DATAQ Instruments, Akron,

Ohio) and Windaq software at a sampling rate of 20, 000 Hz per channel.

For the microphone and airflow signals, the first cursor vertical line was placed at the location of the burst signifying the release of stop consonant closure. When accurate determination of the burst location in the microphone signal was difficult to visualize, coupled views of the wideband airflow and microphone signal were utilized. The airflow signal was used to substantiate the location of the burst in the microphone signal as the release of stop closure is marked in the airflow signal by an overt change in slope and is thus more sensitive to burst detection (**Figure 1**).

Figure. 1: Utility of the Wideband Airflow Signal for

Wideband Airflow

Microphone Signal
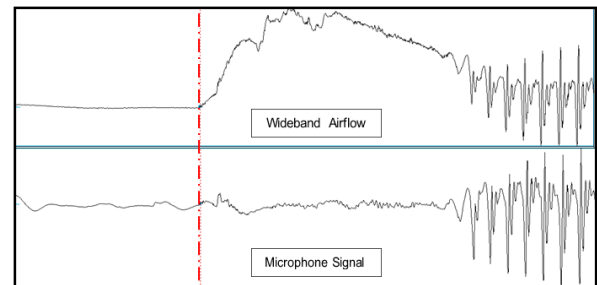
Burst Detection.

A second cursor vertical line was placed at the point of initial detection of phonation (IDP) in the succeeding vowel. The initial detection of phonation was demarcated as the **first valley** of the first regularly occurring signal oscillation that was related in an obvious way to phonation (**Figure 2**). The time elapsed between the two vertical cursors was taken as the VOT. Due to varying measurement methodologies found in the VOT research literature, additional VOT measurement values were also determined using the ***first peak***, ***second valley***, ***second peak***, and ***third valley*** as the first initial detection of phonation in the oscillation of the wideband airflow and microphone signals.

For the wideband spectrogram the initial detection of phonation was measured as the first distinct voicing striation in the first and second formant, which appears to be the standard measurement when using the spectrogram (**Figure 2**).

For the electroglottograph signal, the initial detection of phonation was measured at the first glottal pulse for the initiation of oscillation indicating the initiation of vocal fold contact (**Figure 2**).
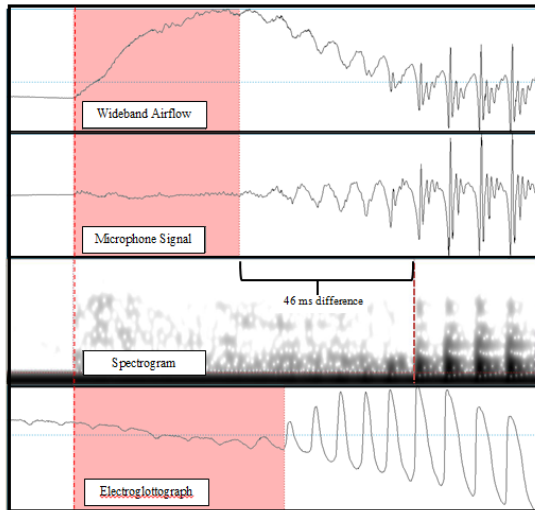
Figure. 2: Segmentation between the burst and first valley as the initial detection of phonation (IDP) for the determination of VOT.

Statistical Analysis

A linear mixed model analysis was conducted to determine the extent to which the four measurement approaches and the measurement locations affected VOT. To examine the effect of measurement approach, a model was constructed using measurement methodology, voicing characteristic, and the associated interaction effect as fixed effects. To examine the effect of measurement location, a model was constructed using location, measurement approach, and the interaction as fixed effects. The random effects structure for both models included voicing characteristic, place of articulation, and trial as random slope terms, and participant as the random intercept term.

RESULTS

Measurement Methodology

Results of the linear mixed model analysis revealed that the *voiceless* stop consonant VOT measurements significantly differed between measurement approaches. However, the VOT measurements of voiceless stops made using the microphone display at its first valley location were not significantly different from the measurements made using the wideband airflow display at its first valley location, $p>0.05$. The VOT measurements of voiceless stops made using the EGG display at the first glottal pulse were significantly different from the measurements made using the wideband airflow display and its first valley location, $p<0.0001$ (by 8.1 ms longer for the EGG signal). The VOT measurements of voiceless stops made using the wideband spectrogram and its first formant excitation were significantly different from the measurements made using the wideband airflow display and its first valley location, $p <0.0001$ (by 13.6 ms longer for the spectrogram).

The only difference between measurement approaches observed for the *voiced* stop VOT measures was for the measures made between the wideband flow or microphone

signals to the spectrographic display. The VOT measures of voiced stops made using the wideband spectrogram at its first formant excitation were significantly different from the measurements made using the wideband airflow display and its first valley location, $p = 0.007$ (by 6.3 ms).

Measurement Location

Compared to the airflow first valley location, all other measurement locations were significantly longer, p<0.05 for all contrasts.

DISCUSSION AND CONCLUSIONS

Results of this study indicate the following:

(1) The use of the initial detection of phonation (IDP) resulted in similar measures of VOT for the wideband airflow and microphone signals, but much shorter intervals than for the electroglottographic signal (about 7.2 ms shorter) and for the spectrographic display of excitation of the formants (about 12.5 ms shorter).

(2) The stop consonant release burst is easy to identify in the wideband airflow signal and is more consistently identifiable in comparison to the microphone signal.

(3) The use of both the microphone and the wideband airflow signals simultaneously reinforce each other for the detection of the articulatory burst as well as the initial detection of phonation, since, for the latter, the glottal oscillation may be subtle during the first cycles, but simultaneous and present on both the airflow and microphone signals, especially when attempting to discriminate glottal oscillation effects from random noise or non-glottal effects.

Future research should consider the use of high-speed video to verify the vocal fold motion simultaneity with airflow and microphone variation that is related to the initial detection of phonation. Additionally, the current study only examined syllable initial productions preceding the /ɑ/ vowel and VOT is known to vary in regards to linguistic task and vowel height; thus, future research should examine differences in measurement approaches in a variety of linguistic contexts.

REFERENCES

[1] Koenig, L. Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research* 2000; 43:1211-1228.

INTRAGLOTTAL PRESSURES FOR GLOTTAL POLYPS

Leila Rahiminejad Ranjbar[1], Abdollah. A. Afjeh[1], Ronald Scherer[2]

[1] Mechanical, Industrial & Manufacturing Engineering Department, University of Toledo, Toledo, OH, U.S.A.
[2] Department of Communication Sciences and Disorders, Bowling Green State University, Bowling Green, OH, U.S.A.

Keywords: Aerodynamics; Intraglottal Pressures; Phonation; Polyps

INTRODUCTION

The vibration of the vocal folds depends upon the airflow and air pressure from the lungs and the recoil of the vocal fold tissue. Therefore, understanding airflow and air pressure is a key part of how sound is generated and altered.

Vocal lesions often change normal vocal fold motion and vocal quality toward a rough or breathy voice, but there are few studies of the aerodynamics of vocal lesions [1,2]. The goal of this study was to obtain air pressure distributions on the surface of the vocal folds in the presence of vocal polyps using laryngeal model M6 [3,4].

The glottis takes on three basic shapes during the phonation cycle -- a convergent shape during opening, an approximately uniform shape near maximum excursion, and a divergent shape when closing. While the glottal shape changes from convergent to divergent, the driving forces change continuously and this promotes vocal fold oscillation. Precise measurement of the forces is the key to understanding basic concepts of vocal fold oscillation, the effects of vocal lesions on vocal fold oscillation, and is basic to understanding voicing aeroacoustics, as well as having empirical data to test computational approaches.

METHODS

The vocal fold model used for this research was M6, an eccentric, three-dimensional physical model of the larynx. M6 is 7.5 times larger than life-size (LS) and has a symmetrical sinusoidal glottal surface profile. The glottis has a maximum width of 0.16 cm (LS) at the mid-coronal section. To capture the effects of polyp lesions, two sessile polyp-like "growths" - spherical caps - of height 0.06 cm and 0.14 cm (LS) were placed unilaterally. There are three rows of 14 pressure taps, located in the inferior to superior parts on the vocal fold surface at locations of the anterior (1/4), middle (1/2), and posterior (3/4) of the anterior-posterior span (1.2 cm LS). Air pressure distributions were obtained for three glottal angles (10° convergent, 0° uniform, 10° divergent), three polyp conditions (no polyp, the protrusion of 0.06 cm, and the protrusion of 0.14 cm) and a range of transglottal pressures (i.e., 1, 5, 15, and 25 cm H_2O LS). The lesions were placed in the medial center of the vocal fold surface (anteriorly-posteriorly and inferiorly-superiorly, Fig. 1). All the cases were considered with the presence of the arytenoid cartilage posteriorly (with no airflow through the posterior glottis). A 48-channel Scanivalve pressure scanner was used. The commercial CFD solver used was ANSYS Fluent 17.0 for laminar flow.
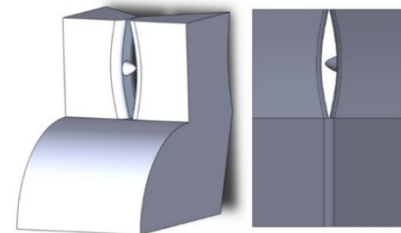

Fig. 1: 0.14 cm height polyp placed on the surface of the vocal folds.

RESULTS

Although some separation occurs in some points and the flow starts to transit to turbulence downstream of the glottis, still the laminar flow model is a good match in the numerical studies for the normal glottis. It is seen in Fig. 2 that the presence of the smaller and larger polyps cause a pressure elevation on the opposite wall, especially upstream of the narrowest location (at zero of the horizontal axis). Fig. 3 suggests that the smaller polyp has relatively little effect on the pressures on the opposite wall, but some effect on the wall with the polyp (via the Fluent trace).
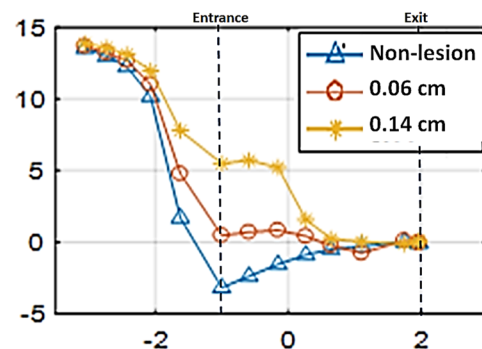

Fig. 2: Pressure data from model M6 on the wall opposite the polyp, divergent glottis, 15 cm H_2O transglottal pressure; distributions for glottis with 0.00, 0.06, and 0.14 cm size polyps. Zero on the horizontal axis is the location of the smallest diameter (tip of the polyp).
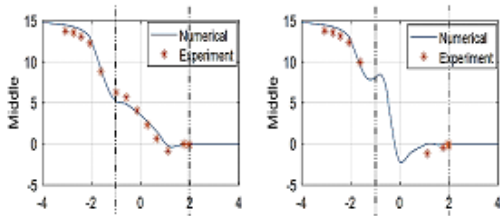
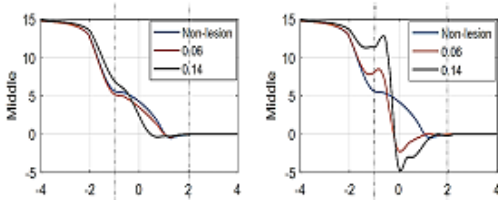Fig. 3. Convergent glottis, pressure distributions for 15 cm H2O transglottal pressure, 0.06 cm lesion.

Fig. 4. Convergent glottis, pressure distributions for 15 cm H2O transglottal pressure, 0, 0.06, 0.14 cm lesion
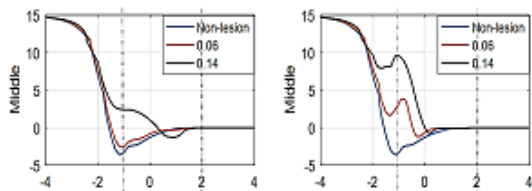

Fig. 5. Divergent glottis, pressure distributions for 15 cm H2O transglottal pressure, 0, 0.06, 0.14 cm lesion
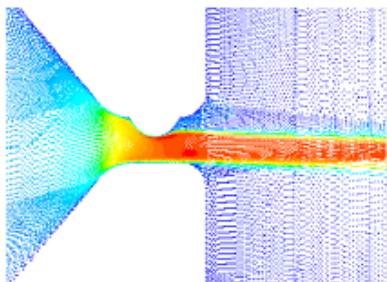

Fig. 6. Velocity contour - Midcoronal view of the uniform glottis.

Figures 4 and 5 indicate that for the convergent and divergent glottis, the larger lesion creates a relatively large pressure driving the polyp itself (right side of the figures), and also alters the pressures on the opposite wall, increasing the pressures more for the larger lesion. These pressures are at the mid-section of the glottis; the pressure effects away from the lesion are much reduced, and follow more normal patterns of pressure distributions found when the vocal folds do not have lesions.

Figure 6 shows the velocity contour for a uniform glottis case at the mid-section of the glottis, showing that the polyp has a significant effect on the flow tube at that location. Indeed, the lesion also creates a double-flow duct with the lesion as the bifurcator.

CONCLUSION

In the glottis for all three angles (convergent, divergent, and uniform geometry), the presence of unilateral sinusoidal (spherical cap) polyps changes the pressure distribution on the surface of the vocal folds of the physical model M6. On the lesion side, middle row, the pressure distributions create relatively very high pressures on the polyp on the upstream side, and then quickly goes negative at the location of the least diameter (at zero on the figures), with subsequent return to the downstream pressure. And on the non-lesion side, middle row, the pressures are all increased especially on the location upstream of the lesion. Thus, the presence of the polyp creates a local driving pressure function that attempts to move the vocal folds laterally despite the glottal angle, suggesting a disruption of the vocal fold motion, especially since anterior and posterior to the polyp, the pressure distribution is less affected. This leads to a bi-directional pressure distribution along the vocal folds, potentially exciting a variety of modes of oscillation, together with the bifurcated glottal flow on either side of the polyp.

REFERENCES

[1] Scherer, R.C., Guo, C.G. (1997). "Intraglottal pressure profiles for normal and abnormal glottal shapes", First International Conference on Voice Physiology and Biomechanics, Evanston, IL, 29 May 97 - 1 Jun 97.

[2] Zhang, Y., Jiang, J.J. (2004). "Chaotic vibrations of a vocal fold model with a unilateral polyp," J. Acoust. Soc. Am 115, 1266-1269.

[3] Torkaman, S. (2011). *Experimental and Computational Study of Intraglottal Pressures in a Three-Dimensional Model with a Non-Rectangular Glottal Shape*. Dissertation, The University of Toledo.

[4] Scherer, RC, Torkaman, S., Kucinschi, B.R., Afjeh, A.A. (2010). "Intraglottal pressures in a three-dimensional model with a non-rectangular glottal shape," J. Acoust. Soc. Am. 128/2, 828-838.

GEOMETRIC MORPHOMETRICS APPLIED TO LARYNGEAL FUNCTIONAL MORPHOLOGY.

Tobias Riede[1], Heather Borgard[1], Karen Baab[1], Bret Pasch[2]

[1] Department of Physiology, Midwestern University, Glendale, AZ, USA
[2] Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011

***Keywords*: Laryngeal function, Morphometrics**

Abstract:

Vocal organ morphology represents an important source of acoustic variation. The absence of homologous landmarks poses a major challenge to comparative analyses. Here we use a geometric morphometric approach based on non-homologous landmarks to quantify shape variation of laryngeal cartilages of four rodent species representing three families. Cartilages of the larynx were reconstructed from contrast-enhanced micro-CT images. Well-established geometric morphometric techniques were used to acquire and superimpose 3D landmarks from the four laryngeal cartilages, and to visualize shape variation. Pairwise Procrustes distance calculation identified minimum numbers of landmarks required in order to sufficiently explain shape variation. For shape analysis, landmark sets were superimposed using generalized Procrustes analysis. Standard multivariate statistics, including principal components analysis, were used to assess taxonomic variation in shape. We found that the four species occupy distinct positions on principle component morphospace, with variation explained in part by phylogeny, body size, and vocal function. Our findings provide the foundation for quantifying the contribution of vocal organ morphology to acoustic diversification.

DEVELOPMENT AND ANALYSIS OF 3D-PRINTED SYNTHETIC VOCAL FOLD MODELS

Ryan G. T. Romero, Taylor E. Greenwood, Clayton A. Young, Serah Hatch, Mark B. Colton, Scott L. Thomson

Department of Mechanical Engineering, Brigham Young University, Provo, Utah, U.S.A.

Keywords: Synthetic Vocal Fold Models; Ultra-soft 3D Printing; Freeform Reversible Embedding

INTRODUCTION

Synthetic self-oscillating vocal fold (VF) models fabricated using silicone materials have been used in voice research to study the flow-induced vibration of models with various geometric and material stiffness characteristics [1]. The advantage of such synthetic VF models over static and driven models is the inclusion of life-like fluid-structure-acoustic interaction effects.

One of the challenges in the development of synthetic VF models is that the current production method (i.e., casting and molding) is relatively inefficient. Modifying the geometry of a VF model using the current method is time-consuming since it requires the design and fabrication of new positive and negative molds. Further, modulus values within the VF are found in discrete areas (layers) within the model. Increasing layer complexity increases the difficulty of, and likelihood of failure during, production.

To address some of these challenges, a custom process for creating 3D-printed synthetic VF models that accurately replicate the form and function of current models is being developed. To date, prototype silicone VF models have been 3D-printed that show potential for 3D printing to be used to fabricate VF models. In the following sections, the 3D printing method is described and vibratory characteristics using the prototypes are presented.

METHODS

The 3D printing process for creating synthetic VFs is based on the method of freeform reversible embedding (FRE) [2,3]. FRE is a form of 3D printing designed to enable printing of a low viscosity material by extruding into a support matrix that holds the material in place until it is cured. FRE printing overcomes the challenges encountered when printing silicone with a relatively low viscosity and long curing time. The FRE process in [2,3] was modified by [4,5] to introduce a newly developed support matrix named organogel [4,5]. In the present study, the process is being further modified to meet the geometric and material property requirements for VF modeling.

Model Generation

Three-dimensional computer models of VF geometry were designed using CAD software (SolidWorks, solidworks.com) and exported as stereolithography (.stl) files. Using these files, g-code (numerical control language) was generated using Slic3r (slic3r.org), an open-source g-code generator.

Printer Design & Parameters

Models were printed using a three-axis CNC mill (Zen Toolworks 12 X 12 F8, zentoolworks.com) controlled by Mach3 CNC Control Software (machsupport.com). A custom syringe extruder mounted to the CNC mill used a linear stepper motor (Haydon Switch & Instrument, Inc.) to extrude material out of a 10 mL syringe with a blunt-tip needle (Howard Electronic Instruments, Inc.)

Printing parameters such as travel speed, layer height, and path width are functions of the needle diameter, silicone viscosity, and support matrix viscosity. Suitable parameter ranges were found using equations found in [4,5]. The specific parameters used during printing were then found using empirical experimentation.

Material Preparation

Traditional synthetic VF models have been fabricated using addition-cure Ecoflex 00-30 (Smooth-On, Inc.) In this study, UV-cure silicone (Momentive Silopren® UV Electro 225-1) was used because of its favorable mechanical properties and because it only polymerizes upon exposure to UV light. The Momentive base material was thoroughly mixed with its catalyst at a 10:1 base:catalyst weight ratio. To lower both the uncured viscosity and cured stiffness of the silicone, the base and catalyst mixture was mixed with Silicone Thinner (Smooth-On, Inc.) at a 1:3 base:thinner weight ratio. The final mixture was then degassed and carefully loaded into the syringe to ensure no air inclusions.

The organogel support matrix was prepared according to instructions outlined in [4], with 2.25 wt% of each polymer specified therein. The organogel was then deposited into small clear containers for printing of individual parts.

Post-Processing of Printed Parts

After printing, the container of organogel holding the completed print was cured in a 350-400nm curing bed (Polylux 2000) for 6 minutes. After curing, the print was removed from the organogel and cleaned with acetone. An epithelial layer was then applied by pouring a thin layer of silicone (Dragon Skin 1A:1B by weight; Smooth-On, Inc.) over the print as described in [6] and allowing to cure at room temperature for 2 hours.

Analysis

Printing parameters such as layer height, print speed, extrusion flow, and path width each affect the final print quality. To tune these parameters, series of single and multi-layer test prints were carried out and analyzed for their overall print quality. Once the desired layer characteristics were achieved, full VF model prototypes were printed and analyzed.

RESULTS AND DISCUSSION

While the 3D printing process is still being developed, the geometry and flow-induced vibratory characteristics of preliminary 3D-printed VF model prototypes have been tested and found to compare favorably with published data from synthetic and excised larynx models, as well as with in vivo data [6]. Images of a VF model during printing are shown in Fig. 1. Prototype prints were produced using a 25-gauge needle with a print speed of 5 mm per second, and a layer height of 450 microns. Printing and post-processing time was approximately 1.5 hours. The shape and size of 3D-printed prototype VF models closely match the desired geometry (see Fig 2).

The 3D-printed models exhibited flow-induced vibration when mounted in the test setup (see Fig. 3). The vibration frequency of the model shown in Fig. 3 was 110 Hz, within the human vocal range [6]. The onset and offset pressures ranged between 2.23 and 2.56 kPa, and 2.03 and 2.16 kPa, respectively.
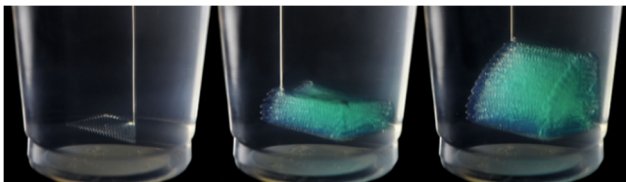

Fig 1: VF model being 3D-printed at the start (left), an intermediate stage (middle), and at completion (right).
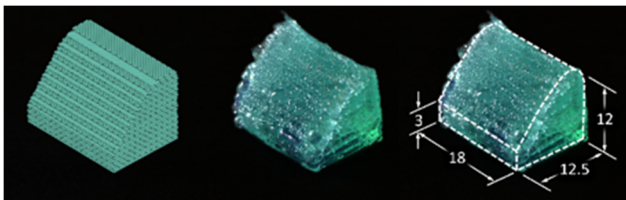

Fig 2: Isometric view of a VF model rendered in Slic3r (left), fully printed and cured (middle), and dimensioned with edges highlighted (right, [mm]).
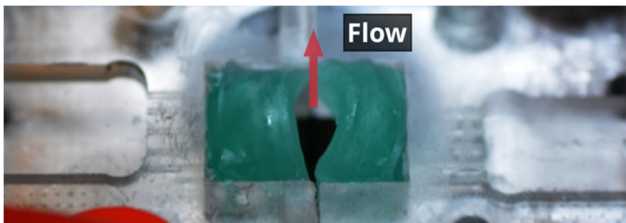

Fig 3: 3D-printed VF models during vibration.

CONCLUSION

These preliminary results show that 3D printing of phono-mimetic synthetic vocal fold models has potential as a viable and agile fabrication alternative over the current process. The advantages of 3D printing facilitate geometry changes, which allows fine-tuning of geometric features with reduced manufacturing overhead. Further refinement is needed, however, and research will focus on improving print resolution and geometric accuracy. It is anticipated that the 3D printing process will accelerate the development of VF models and facilitate improved subsequent exploration of the influence of geometry and other key parameters on VF vibration.

ACKNOWLEDGMENTS

REFERENCES

[1] Kniesburges S., Thomson S. L., Barney A., Triep M., Šidlof P., Horáček J., Brücker C., and Becker S. (2011). "In vitro experimental investigation of voice production," Current Bioinformatics 6(3):305–22.

[2] Hinton, T. J., Hudson, A., Pusch, K., Lee, A., and Feinberg, A.W. (2016). "3D printing PDMS elastomer in a hydrophilic support bath via freeform reversible embedding," ACS Biomaterials Science and Engineering 2(10):1781–86.

[3] Hinton, T. J., Jallerat, Q., Palchesko, R. N., Park, J. H., Grodzicki, M. S., Shue, H., Ramadan, M. H., Hudson, A. R., and Feinberg, A. W. (2015). "Three-dimensional printing of complex biological structures by freeform reversible embedding of suspended hydrogels," Science Advances 1(9).

[4] O'Bryan, C. S., Bhattacharjee, T., Hart, S., Kabb, C. P., Schulze, K.D., Chilakala, I., Sumerlin, B.S., Sawyer W. B., and Angelini, T. E. (2017) "Self-assembled micro-organogels for 3D printing silicone structures," Science Advances 3(5).

[5] O'Bryan, C. S., Bhattacharjee, T., Niemi, S. R., Balachandar, S., Baldwin, N., Ellison, S. T., Taylor, C. R., Sawyer, W. G., and Angelini, T. E. (2017). "Three-dimensional printing with sacrificial materials for soft matter manufacturing," MRS Bulletin 42(8):571–77.

[6] Murray, P. R., Thomson, S. L., (2011) "Synthetic, multi-layer, self-oscillating vocal fold model fabrication," Journal of Visualized Experiments (58):3498.

EFFECT OF METHYLPREDNISOLONE TREATMENT ON RABBIT VOCAL FOLD PHYSIOLOGY

Lea Sayce, DPhil[1]; Emily E. Kimball, MS[2]; Gary Gartling, MS[2]; Maria Powell, PhD, CCC-SLP[1]; Shintaro Sueyoshi, MD[1]; Steven Schneeberger, BS[3]; Jennifer Brandley, BA[2]; Carol Xu, MD, PhD[1]; Bernard Rousseau, PhD, MMHC, CCC-SLP[1,2,4]

[1] Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, Tennessee, USA.
[2] Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, Tennessee, USA.
[3] School of Medicine, Vanderbilt University, Nashville, Tennessee, USA.
[4] Department of Mechanical Engineering, Vanderbilt University, Nashville, Tennessee, USA.

Keywords: Glucocorticoid steroids; Immunofluorescence; Vocal folds; Safety.

INTRODUCTION

Dysphonia is a common condition, affecting an estimated $20 - 23$ million Americans annually [1], of which around 18% of cases are caused by phonotraumatic vocal fold lesions [2]. Glucocorticoid steroids are frequently administered for the treatment of phonotraumatic vocal fold injuries [3], and a small population of retrospective and prospective case studies suggests that treatment with steroids may lead to decreased lesion size [4], lesion resolution [5], symptomatic improvement [5, 6], and the avoidance of phonomicrosurgery [7]. However, despite being widely used, our understanding of the mechanism of action, dose-response, safety, and efficacy of glucocorticoid steroid treatments for vocal fold injury is extremely limited. This study utilizes a rabbit model to investigate the effects of methylprednisolone on vocal fold physiology.

METHODS

Glucocorticoid Steroid Treatment

This study was approved by Vanderbilt IACUC under protocol M1600081. Ten age and weight matched New Zealand white rabbits were systemically dosed with 4.5mg of methylprednisolone (n=5) or volume matched saline (n = 5) for 6 consecutive days. Following a 14-day recovery period the rabbits underwent a non-survival tissue harvest procedure. Briefly, rabbits were anesthetized and shaved, an incision was made from sternum to submentum, and the underlying fascia and muscle were dissected to expose the larynges. The animal was then euthanized and the larynx was excised. One half of the larynx was fresh-frozen in Optimal Cutting Temperature (O.C.T) medium, while the other half was subdissected to isolate the vocal fold, which was subsequently stored in RNAlater for transcript analysis. Weights were recorded at baseline, daily throughout the treatment course, and prior to terminal surgery.

Immunolabeling

Fresh-frozen tissue was sectioned using a Leica CM1800 cryostat at 12-micron thickness. Slides were labeled using primary antibodies against junctional proteins zonula occludin-1 (ZO-1, Invitrogen 3391000, 1:200) and e-cadherin (BD Transduction 610181, 1:100), markers of vascularity ZO-1 and CD31 (Novus Biologicals NB600-562, 1:100), and intracellular ion channels NaK$^+$ATPase (Abcam ab7671, 1:100) and Aquaporin 1 (Santa Cruz sc25287, 1:100). Samples were labeled with Goat-anti-Mouse Alexafluor 594 (Invitrogen, 1:200) and co-stained with DAPI. Images were captured using a Nikon Eclipse 90i microscope with a 20x objective and analyzed using ImageJ.

Quantitative PCR

Vocal folds were homogenized in TRIzol and RNA was extracted. cDNA was synthesized via reverse transcription and transcript levels were detected using quantitative polymerase chain reaction (qPCR). Custom 5'FAM, 3'TAMRA dual-labeled probes and primers were designed to target the following genes of interest: *E-cadherin*, *tight junctional protein-1 (tjp-1,* gene encoding ZO-1), *Cyclooxygenase-2* (*Cox-2*), and transforming growth factor – beta (*tgf-β*). Relative expression levels ($\Delta\Delta C_T$) were calculated using succinate dehydrogenase complex subunit A (SDHA) as an endogenous control. Reactions were performed in triplicate for 40 cycles with an annealing/extension temperature of 60°C. StepOnePlus Software v2.3 was used to calculate $\Delta\Delta C_T$.

Analysis

Immunolabeling: Protein localization within the vocal fold was assessed manually. Protein expression was semi-quantitatively analyzed for intensity using Otsu thresholding and Analyze-Measure tools on ImageJ, while protein area was analyzed using Otsu thresholding, followed by the Analyze Particles tool on Image J. Protein intensity and protein area across treatments were analyzed by Mann-Whitney U test. qPCR: Levene's test, followed by a two-tailed Student's t-test was performed for each assay to assess intra-treatment differences in transcript levels. All analyses were performed in SPSS v24.0.
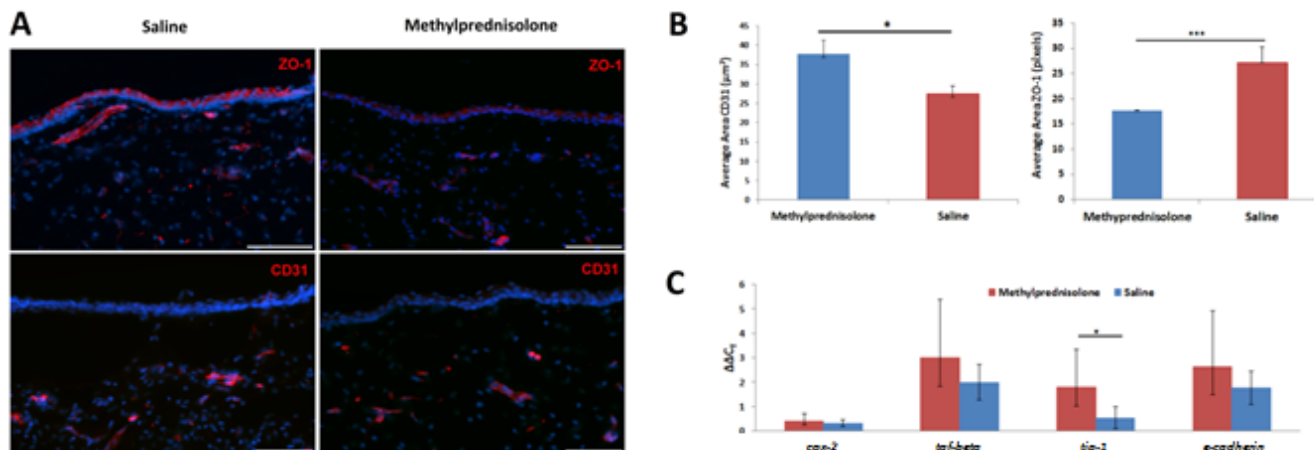
Fig. 1. Methylprednisolone induces differential expression of endothelial ZO-1 and CD31 in the rabbit vocal fold. A and B. CD31 area is increased and ZO-1 area is decreased in the vocal folds of methylprednisolone-treated rabbits. **C**. qPCR demonstrated significant downregulation of *tjp-1* in vocal fold tissue of methylprednisolone-treated rabbits.

RESULTS

Methylprednisolone-treated animals demonstrated significant relative weight loss during the 14 day recovery period (p=0.0002), consistent with withdrawal from steroids. Immunolabeling for CD31 revealed larger vessels, with a reduced average area of ZO-1 endothelial expression, in the vocal fold lamina propria of methylprednisolone-treated animals vs. saline-treated controls (p=0.04 and p=8x10-5 respectively). No changes in protein localization or semi-quantitative protein expression were observed in the vocal fold epithelium. qPCR analyses identified down-regulated *tjp-1* expression (p= 0.037) in methylprednisolone-treated animals, vs. saline-treated controls. No differences were identified between treatment conditions for *cox-2*, *tgf-β*, or e-*cadherin*.

DISCUSSION AND CONCLUSIONS

This study identifies that systemic treatment with the glucocorticoid steroid methylprednisolone elicits long-term changes to the vasculature of the lamina propria. The average area of the vasculature was increased in methylprednisolone-treated animals, while the key tight-junctional protein ZO-1 was found to have reduced area. Additionally, at the transcript level *tjp-1* was downregulated. While not investigated in this study, the vascular changes observed may present an increased risk of vocal fold hemorrhage through enlarged size and decreased luminal barrier integrity. This study has several limitations; the steroid dose (4.5mg) was extrapolated from a dose that is 3 times higher than typically administered in humans, thus the dosing schedule may not be representative of clinical practice. While humans are a steroid hyposensitive species, rabbits are steroid hypersensitive [8]. As such, data obtained in this study may reflect a more overt pathology than expected in humans. Further studies are required to elucidate the timeline,

permanence, dose-response, and mechanism of actionthrough which methylprednisolone effects these changes to the vocal folds.

REFERENCES

[1] Cohen SM, Kim J, Roy N, Asche C, Courey M. Direct health care costs of laryngeal diseases and disor-ders. *The Laryngoscope.* 2012;122(7):1582-1588.

[2] Cohen SM, Kim J, Roy N, Asche C, Courey M. Prevalence and causes of dysphonia in a large treatment-seeking population. *The Laryngoscope*. 2012;122(2):343-348.

[3] Rafii B, Sridharan S, Taliercio S, et al. Glucocorticoids in laryngology: a review. *The Laryngoscope*. 2014;124(7):1668-1673.

[4] Wang C-T, Liao L-J, Cheng P-W, Lo W-C, Lai M-S. Intralesional steroid injection for benign vocal fold dis-orders: A systematic review and meta-analysis. The Laryngoscope.2013;123(1):197-203.

[5] Lee S-H, Yeo J-O, Choi J-I, et al. Local steroid injection via the cricothyroid membrane in patients with a vocal nodule. Arch Otolaryngol Head Neck Surg. 2011;137(10):1011-1016.

[6] Lee S-H, Yeo J-O, Choi J-I, et al. Local steroid injection via the cricothyroid membrane in patients with a vocal nodule. Arch Otolaryngol Head Neck Surg. 2011;137(10):1011-1016.

[7] Mortensen M, Woo P. Office steroid injections of the larynx. The Laryngoscope. 2006;116(10):1735-1739.

[8] Cohn, L. A. (n.d.). The Influence of Corticosteroids on Host Defense Mechanisms. Journal of Veterinary Internal Medicine, 5(2), 95–104.

INFLUENCE OF SPATIAL CAMERA RESOLUTION ON GLOTTAL AREA WAVEFORM PARAMETERS IN HSV IMAGING

Patrick Schlegel[1], Melda Kunduk[2], Michael Döllinger[1]

[1] Dep. of Otorhinolaryngology, Div. of Phoniatrics and Pediatric Audiology, University Hospital Erlangen, Erlangen, Germany
[2] Dep. of Communication Sciences and Disorders, Louisiana State University, Baton Rouge, LA, USA

Keywords: **High-Speed Videoendoscopy; Glottal Area Waveform; Camera Resolution; Parameters**

INTRODUCTION

High-speed videoendoscopy (HSV) is a powerful tool for investigating vocal fold vibratory function. However, like any other method of measurement, also HSV is prone to a wide spectrum of different factors that are expected to influence HSV parameters. These factors include the recording frame rate, the analyzed interval length and the spatial image resolution. Although the influence of the recording frame rate was already investigated to some extend [1] the effect of the spatial resolution still remains widely unknown.

From HSV images, the glottal area can be calculated over time or frames. The resulting signal is referred to as glottal area waveform (GAW). Different definitions of the GAW exist [2-4]. In this work, the GAW is defined as the simple function of the glottal area in pixels over frames. All parameters that were investigated in this work are based on this definition of the GAW. Depending on the camera resolution the recorded images and consequently the GAW have a different level of accuracy.

In total we investigated the influence of varying image resolution on 50 different GAW parameters. The aim of this work is to demonstrate the sensitivity of different types of parameters to the changes in the image resolution.

METHODS

In total 40 HSV recordings of 20 healthy male and 20 healthy female subjects during sustained phonation were investigated. A Photron Fastcam MC2 camera with a frame rate of 4000 fps and (1) a spatial resolution of 512x256 pixels was used. This initial resolution was reduced by pixel averaging to (2) a resolution of 256x128 pixels and (3) to a resolution of 128x64 pixels, yielding 40 sets of recording triplets. The original recordings were selected from a larger pool of HSV videos and have a high recording quality with its brightness and contrast which ensured the recordings after resolution reduction, were segmentable without artifacts being introduced. However, the distance between endoscope camera and glottis and furthermore the sizes of larynges of the subjects vary between the recordings. Hence, the pixel area on the recordings that is taken up by the vocal folds and therefore the glottal area also differ. For men the global GAW maximum fluctuates between 742 and 3214 pixels among the different original recordings, for women the range is between 978 and 2057 pixels.

Every recording triplet was segmented using a modified in house developed version of Glottis Analysis Tools (GAT – 2018). The procedure was as follows:

1. A 1000 frames segment of the original recording was selected.
2. In the recording with the lowest resolution the glottal area was covered with a regularly mesh of seed points.
3. This mesh was condensed by the factor 2 and 4 respectively for the second lowest and highest resolution.
4. For the highest resolution the brightness-thresholds were adjusted.
5. The same thresholds were assigned for the other two resolutions.

Fig. 1 illustrates the condensing of the seed points between the different resolutions of one triplet of recordings. From the lowest resolution (left) to the original resolution (right) the number of seed points increases whilst their relative density per pixel stays the same (one seed point per two pixels in the mesh region). The mesh itself covers the main part of the glottis especially the area where the vocal folds collide and therefore the glottal area narrows.

For each seed point all surrounding pixels with a lower brightness value than the chosen threshold are selected (bright gray area in Fig. 1) until the first pixels of an equal or higher brightness are reached. By these means darker regions on the 2D map of the image, like the glottis opening, can be marked. Islands of brighter pixels surrounded by marked pixels are also marked. The brightness-thresholds are set to match the relative brightness of the image in the horizontal plane. For example in Fig. 1, the thresholds decrease from top down. As a consequence, in the upper section of the image brighter pixels are included to the glottal area than in the lower section of the image.

For each triplet of recordings 1000 frames are segmented per recording yielding a triplet of GAWs. For every GAW 50 different cycle-based parameters are computed. These parameters are determined for the same

20 max-based cycles for every triplet of GAWs resulting in 40 triplets of 50 different parameters. Parameters that were calculated for single cycles were averaged over all 20 cycles.
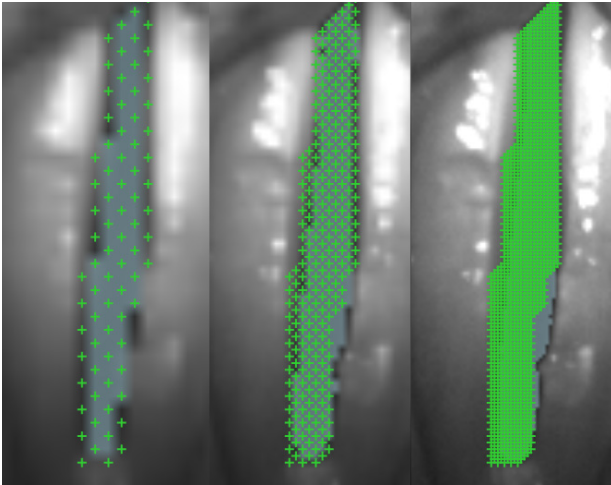

Fig. 1: One triplet of recordings from lowest resolution (left) to original resolution (right). The bright grey area illustrates the segmented section of the glottal gap. Each green plus marks one seedpoint.

RESULTS AND DISCUSSION

Measures indicating the mean fundamental period like fundamental frequency or mean cycle duration showed almost no deviation within the recording triplets. In contrast, measures for computing period perturbation like mean jitter or period perturbation factor displayed deviations of 100% and more for some subjects, for other subjects almost no deviation occurred. This was also the case, if the initial jitter values were not very close to zero. This difference is based on the limited change in total signal length versus a possible change in length of each cycle due to the shift of cycle maxima.

Similar to jitter measures shimmer measures changed for lower resolutions, as well, although to a greater extend. While the degree of jitter declined for some recording triplets from high to low resolutions and rose for others, the degree of shimmer was larger for lower resolutions in almost every subject. This can be explained by a decreasing pixel number of the images and hence a greater relative difference between the images if one pixel more or less is included in the GAW. Energy perturbation measures like energy perturbation factor or different energy perturbation quotients showed deviations within the recording triplets similar to jitter and shimmer measures. This is expectable since energy perturbation is dependent on period and amplitude perturbation by being calculated from the absolute square sum of each cycle.

The degree of symmetry between the left and right GAW slightly declined with decreasing image resolution resulting in a deterioration of most symmetry parameters

compared to the original resolution. Additionally, some GAW parameters on the phase characteristics like open quotient (OQ) and glottal gap index (GGI) reduced implicating an extended closed phase. Parameters connected directly to camera resolution like maximum area declination rate or amplitude–to–length-ratio reduced greatly with decreasing image resolution and pixel number.

The described effects were similar for women and men. However, the reduction of OQ related parameters was more distinctive for male recordings. The effect on GGI was less pronounced in males since in many cases it was already zero in the original recordings due to an existing closed phase and thus could not decrease any further.

The differences in between the recording triplets were relatively large which can be attributed to the large fluctuations in sizes of the glottides in the initial recordings. For this reason and the small number of HSV recordings used in this study, this work should be considered as a first qualitative investigation of the influence of camera resolution.

CONCLUSION

In this work the influence of camera resolution on different objective parameters was shown. However, more studies are needed to investigate this topic further using a larger number of recordings or applying a method for ensuring a constant glottis camera distance.

ACKNOWLEDGMENTS

REFERENCES

[1] Schützenberger A, Kunduk M, Döllinger M, Alexiou C, Dubrovsky D, Smmler M, Seger A, Bohr C. Laryngeal High-Speed Videoendoscopy: Sensitivity of Objective Parameters towards Recording Frame Rate. *BioMed Research International*, vol. 2016 , Article ID 4575437, 19 pages, 2016.

[2] Xin Chen, Diane Bless, and Yuling Yan. A segmentation scheme based on rayleigh distribution model for extracting glottal waveform from high-speed laryngeal images. In *27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS) 2005.*, pages 6269–6272. IEEE, 2006.

[3] Amaia Mendez, Begona Gracia, Ibon Ruiz, and Ibai Iturricha. Glottal area segmentation without initialization using gabor filters. In *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) 2008.*, pages 18–22. IEEE, 2008.

[4] Melda Kunduk, Yuling Yan, Andrew J. McWhorther, and Diane Bless. Investigation of voice initiation and voice offset characteristics with high-speed digital imaging. *Logopedics Phoniatrics Vocology*, 31(3):139–144, 2006.

THE EFFECT OF A POSTERIOR GLOTTAL GAP ON CONTACT PRESSURES IN A SYNTHETIC SELF-OSCILLATING VOCAL FOLD MODEL

Mohsen Motie Shirazi, Byron D. Erath

Department of Mechanical and Aeronautical Engineering, Clarkson University, Potsdam, New York, USA

Keywords: Synthetic Vocal Folds, Glottal Gap, Contact Pressure, Compensation

Abstract:

Introduction and Objectives: The presence of a posterior glottal gap decreases aerodynamic energy transferred to the vocal folds, causing a reduction in the sound pressure level (SPL). Common compensations include increasing the subglottal pressure to boost acoustic output. Higher mechanical stresses resulting from this compensatory behavior are believed to be a primary contributor to the emergence of structural pathologies. The objective of this study is to determine how the presence of a persistent posterior glottal gap influences spatial and temporal variations in vocal fold contact pressure.

Methods: A hemilaryngeal facility is utilized to allow direct measurement of contact pressures. A self-oscillating VF is manufactured from silicone using a four-layer approach and a spatial grid of static pressure taps are placed on the contacting wall. Contact pressure, mean flow rate, subglottal pressure, VF kinematics, and radiated SPL are simultaneously acquired. The influence of three posterior glottal gaps (3, 6, and 9 mm^2) on these measures is investigated.

Results: The inclusion of a glottal gap increases the mean flow rate and onset pressure, and lowers the radiated SPL. Modeling compensatory behaviors by adjusting the subglottal pressure and flow rate to match the radiated SPL of normal speech increases the magnitude of the contact pressures. The impact of compensation on aerodynamic and VF kinematic measures is also discussed.

Conclusions: The inclusion of a posterior glottal gap is shown to have a detrimental effect on VF measures associated with vocal quality, while increasing subglottal pressure to correct for these deficiencies produces pathological levels of contact pressures.

MEDIAL SURFACE PRESSURE DISTRIBUTIONS IN NORMAL MODE BASED GLOTTAL CONFIGURATIONS

Simeon L. Smith[1], Ingo R. Titze[1,2]

[1] National Center for Voice and Speech, University of Utah, Salt Lake City, UT, USA
[2] Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA, USA

***Keywords*: Glottal Pressure; Glottal Shape; Modeling; Voice Simulation**

Abstract:

Objectives: The three-dimensional glottal shape during phonation determines pressure distributions on the medial surface of the vocal folds, which influences vocal fold vibration and the consequent acoustic sound waves. This fluid to structure transfer of energy has been researched extensively, yet glottal pressure distributions are still not well known. Voice modeling requires accurate estimates of pressure for realistic self-oscillation, which is a challenge for complex flow situations that arise in the disordered voice. We set out to quantify glottal pressure distributions using synthetic experimental models, with the eventual goal of improving flow regimes for voice simulation.

Methods: Three-dimensional glottal geometries were derived by applying normal mode displacements to prephonatory vocal fold geometry. Scaled-up polylactic acid (PLA) vocal fold models were 3D-printed for various glottal geometries and secured in a flow tunnel. Air flow was induced through the glottal configuration and pressure distributions on the medial surface were measured via a 7 x 9 array of pressure taps in the 3D-printed models. Medial surface pressure contours (isobars) were calculated from the experimental data.

Results: Medial surface pressure contour plots for several complex glottal shapes are shown. Results showed expected flow behavior for simple flow channel features (pressure drop with reduced area according to the Bernoulli Effect, flow separation in divergent configurations, etc.), along with complex interactions of basic flow phenomena for higher order glottal shapes.

Conclusions: Medial surface pressure distributions were successfully quantified using a scaled-up experimental setup with 3D-printed glottal shapes. These results provide information that will be used to develop flow regimes for voice simulation that are more capable of modeling complex flow associated with the disordered voice.

SPEECH PERFORMANCE DENSITY AS A MEASURE OF LONG-TERM SPEAKING CHARACTERISTICS

Brad H. Story[1], Kate Bunton[1]

[1] Department of Speech, Language, and Hearing Sciences, University of Arizona, Tucson, AZ, USA
Keywords: e.g. fundamental frequency; speech analysis; formants

INTRODUCTION

The acoustic characteristics of speech produced by any given talker may change considerably depending on their physical or emotional state, surrounding environment, and communicative intent. For example, increased distance between talker and listener may result in modifications to sound pressure level (SPL), fundamental frequency (f_o), formant frequencies (F_1, F_2, F_3, …), and segment and pause durations.

The purpose of this study was to develop and evaluate a method for visualizing, quantifying, and comparing the acoustic characteristics of different types of speaking performance both within and across talkers. The word "performance" is operationally defined here as any particular instance of producing speech for the purpose of communicating within a specific context or environment. With an audio recording of at least several minutes of speech as input, the method generates two "speech performance density" plots denoted as SPD_1 and SPD_2 that are derived by extracting the time-varying quantities of fundamental frequency (f_o), sound pressure level (SPL), and the first two formant frequencies (F_1 and F_2). SPD_1 is based on f_o and SPL, where each is normalized to their median values, and f_o converted to semitones, and SPD_2 is based on a normalization of formant frequencies to their median values. The normalization allows for intra- and inter-talker analysis of speaking performance.

METHOD

Human Data

For purposes of illustrating the steps in developing the analysis technique, a recording was obtained of an adult male talker repeating a sequence of 11 sVd words, each containing a different English vowel, over the course of about two minutes. Serving only as a test case, the intent was to assure that the talker spent most of the speaking time producing clear instances of the 11 vowels, while reducing vowel-vowel transitions and eliminating "um", "ah", etc. Additional recordings of the same male talker producing several minutes of speech (i.e., the Caterpillar Passage [1]) in a variety of conditions (e.g., mumble, clear, conversational) were obtained to demonstrate the output of the algorithm. All signals were recording with a sampling frequency of f_s = 44100 Hz.

Initial Processing Steps

1. A periodicity detector was used to find the time points bracketing all voiced segments in an audio file [2]. Subsequent analyses were limited to these segments.
2. Fundamental frequency f_o was obtained with a cycle detection algorithm. The RMS amplitude was determined for 0.025 second windows with an overlap of 0.00625 seconds, and converted to sound pressure level (SPL) in dB.
3. The first two formant frequencies, F_1 and F_2 were estimated at 0.0025 second increments with a peak-picking algorithm applied to an autocorrelation LPC spectrum computed for analysis windows of 0.025 seconds (Gaussian window).

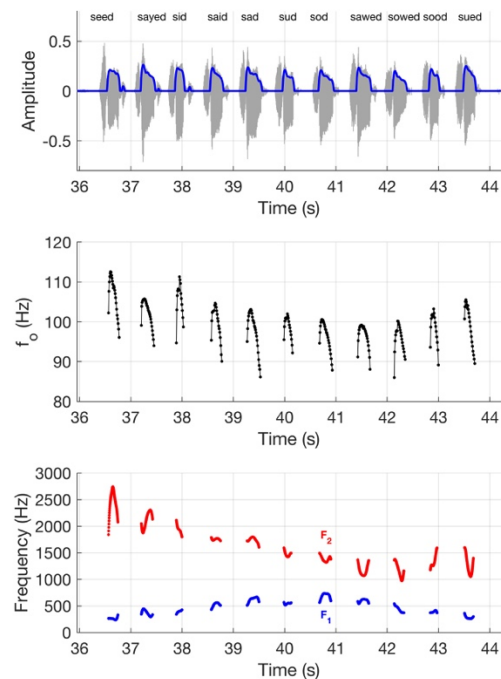These steps are demonstrated in Fig. 1 in the panels ordered from top to bottom.

Fig. 1: Demonstration of initial processing steps for one sVd sequence. Upper – audio signal and RMS (voiced segments only), Middle- f_o, Bottom – F_1 and F_2.

Speech Performance Densities: SPD₁ and SPD₂

SPD₁: SPL and f_o, shown in raw form in Fig. 2a, were normalized relative to their median values, and f_o was converted to semitones. In a grid spanning [-8,12] semitones in the x-dimension and [-8,12] dB in the y-dimension, the number of [f_o, SPL] pairs present within a 0.75 radius of every point in the grid were counted to generate a density value. Density values distributed across the grid show how often the talker was producing f_o and SPL values in a particular region of the SPD_1 space (Fig. 2c). The portion of the SPD_1 space utilized by the talker was quantified by finding the area enclosed by a convex hull at a density level of 0.25 (white curve in Fig. 2c). For this particular case, the repetition of the sVd sequence caused the talker to produce nearly the same change in f_o for each word, thus the SPD₁ is fairly constrained with a convex hull area of only 16.6 dB*semitones.

SPD₂: [F_1, F_2] pairs shown in Fig. 2b were normalized as $(F_n - F_n^{median}) / F_n^{median}$ where n=1, 2. A grid spanning [-1,1] in both the x- and y-dimensions was used to assess the density of the normalized formant pairs in the same manner as SPD_1 and plotted with a color map as described in [3]. The SPD_2 space is a representation showing the talker's distribution of formants over the duration of the recording, and a convex hull area analysis was again used to characterize the density distribution (Fig. 2d). The density plot clearly reveals the locations of the 11 vowels in the sVd sequence. Note that the convex hull area (equal to 0.83 in this case) for SPD₂ is dimensionless.
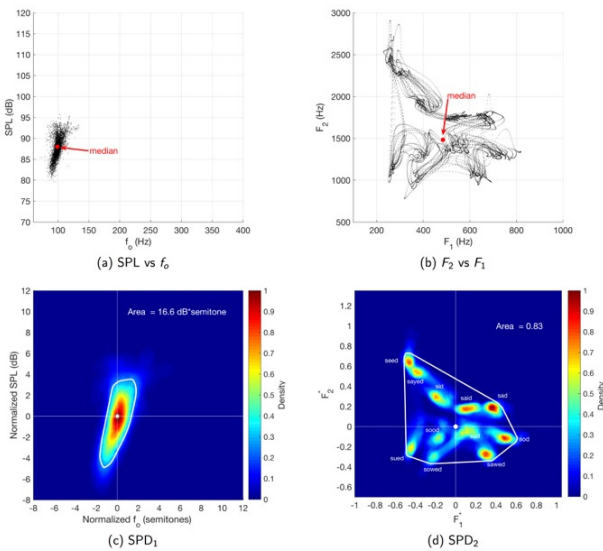

Fig. 2: Transformation of (a) SPL vs f_o and (b) F₂ vs F₁ to normalized (c) SPD₁ and (d) SPD₂.

RESULTS

As a demonstration, the SPD analysis was applied to two recordings of the Caterpillar Passage [1], both spoken by the same adult male. In the first recording, the talker spoke in a mumbly, hypo-articulated manner, whereas, in the second, the talker hyper-articulated in an attempt to produce "clear" speech [4]. The SPD₁ and SPD₂ plots in Fig. 3 indicate small hull areas for the hypo-articulated condition, and large, expansive hull areas for the hyper-articulated case.
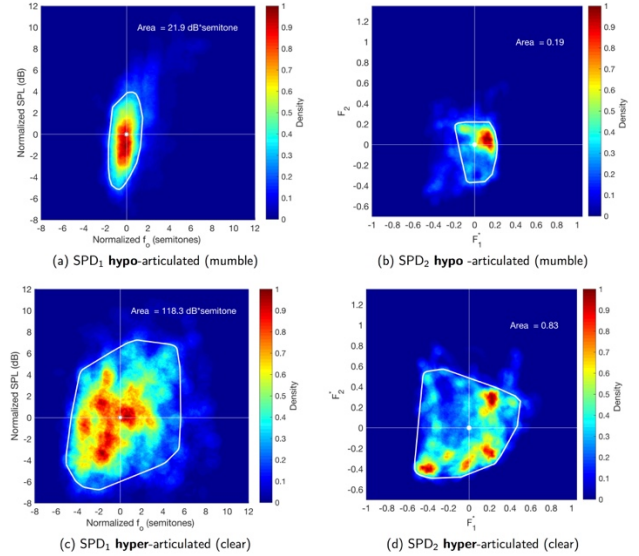

Fig. 3: SPD₁ and SPD₂ plots characterizing hypo-articulated (a & b) and hyper-articulated (c & d) speech.

DISCUSSION AND CONCLUSIONS

The speech performance density plots provide a compact means of visualizing selected acoustic characteristics representing aspects of both voice and speech production. SPD₁ and SPD₂ may be useful for documenting changes in speaking characteristics of a specific talker, or for comparing the same characteristics across talkers.

REFERENCES

[1] Patel R, Connaghan K, Franco D, Edsall E, Forgit D, Olsen L, Ramage L, Tyler E, Russell S. "The Caterpillar": A novel reading passage for assessment of motor speech disorders. Am. J. Spch-Lang Path. 2013 Feb 1;22(1):1-9.

[2] Xie, Z., Niyogi, P. Robust acoustic-based syllable detection. *9th International Conference on Spoken Language Processing*. 2006; 1571-1574.

[3] Story BH, Bunton K. Vowel space density as an indicator of speech performance. J. Acoust. Soc. Am. 2017 May;141(5):EL458-64.

[4] Smiljanić R, Bradlow AR. Speaking and hearing clearly: Talker and listener factors in speaking style changes. Lang ling. comp. 2009;3(1):236-64.

EXPRESSION OF INFLAMMATORY CYTOKINES AND JUNCTIONAL PROTEINS IN RABBIT VOCAL FOLDS FOLLOWING METHYLPREDNISOLONE TREATMENT

Shintaro Sueyoshi , Lea Sayce, Emily Kimball, Gary Gartling, Carol Xu, Bernard Rousseau

Vanderbilt University Medical Center, Nashville, TN, USA

***Key words*: Vocal fold; Epithelium; Rabbit; Human**

Abstract:

Objectives: Glucocorticoid steroids are often prescribed to treat acute voice disorders in general practice. The anti-inflammatory mechanisms of these steroids have been characterized in other organ systems, and level-4 clinical evidence suggests a therapeutic benefit for the treatment of acute voice disorders; however, the foundational molecular biology of steroid administration in the vocal folds has yet to be explored. The objective of this study is to characterize changes to vocal fold gene expression following acute administration of methylprednisolone.

Methods: Ten adult New Zealand white rabbits were randomized to 6 daily intramuscular injections of either 4.5mg methylprednisolone (n= 5), or saline (control, n= 5). Following a 14 day recovery period, animals were sacrificed and the larynges harvested. The vocal folds were sub-dissected and stored in RNAlater at -80o until subsequent processing. Vocal fold tissue was homogenized in TRIzol using a bead beater, and RNA was extracted and used to generate template cDNA. qPCR was performed in triplicate to detect expression of cytokines COX2 and TGF-b, and junctional proteins eCadherin and ZO-1 using custom designed primers and 5'FAM, 3'TAMRA dual-labelled probes. Relative expression levels (DDCT) were calculated using an endogenous control (SDHA), and biological replicates were compared statistically using a two-tailed t-test.

Results: qPCR demonstrated significantly down-regulated ZO-1 expression (p= 0.037) in methylprednisolone treated animals relative to saline controls. No difference in the relative expression of COX2, TGF-b, or eCadherin was detected between groups.

Conclusions: Decreased ZO-1 expression at 14 days post-administration suggests long-term effectson tight junction integrity and epithelial barrier function with acute methylprednisolone treatment. Further studies are required to elucidate the mechanism through which methylprednisolone downregulates expression of this essential epithelial gene.

TAHRIR PATTERNS IN A PROFESSIONAL IRANIAN CLASSICAL SINGER

Mahdi Tahamtan[1] and Ronald Scherer[1]

[1] Communication Sciences and Disorders, Bowling Green State University, Bowling Green, OH, USA

***Keywords*: Classical Singing, Cultural Music, Vocal Ornamentals, Voice**

INTRODUCTION

"Iranian classical singing", called "avaz" in Persian, contains a vocal ornamentation called **tahrir** which is inspired by the sound of the nightingale [1,2]. Tahrir is composed of a continuous chain of tekyehs [3], where each **tekyeh** is a melismatic embellishment [4] with a fundamental frequency gesture that typically quickly increases and then decreases in a very short time interval. As the tahrir is one of the main distinguishing features of avaz from other singing styles, understanding tahrir is necessary in order to understand avaz. Performing artists produce different patterns of tahrir during their singing. The primary aim of the current research is to determine and describe the tahrir patterns produced by a professional avaz singer.

METHODS

A live unaccompanied recording was made of a male professional avaz singer singing the popular and famous song "Morghe Sahar" ("Dawn Bird"). The performer was asked to sing the song in a key of his choosing, but to sing in a manner similar to how "Mohammad Reza Shajarian" has sung the song in the past, the most famous avaz singer of all time. In this study the primary factors found to categorize tahrir patterns include the number of tekyehs, fundamental frequency extent from the baseline to the tekyeh peak, and inter tekyeh interval duration.

RESULTS

For this song and this singer, 4 primary tahrir patterns were identified.

Pattern 1: This tahrir consists of one tekyeh and is called a "zinat". Pattern 1, as a single pulse tahrir pattern occurs on a consonant and could have a prominent downward frequency change before the upward frequency change (type 1A) or on a vowel (type 1B). Type 1A is more frequent than type 1B (Fig. 1). The approximate range of the fundamental frequency extent is 5.36 ST (min of 3.60 ST, max of 8.96 ST).

Pattern 2: This tahrir pattern is a multiple tekyeh gesture with relatively long inter-tekyeh intervals (average of 256 ms). As the plateau intervals are relatively long,
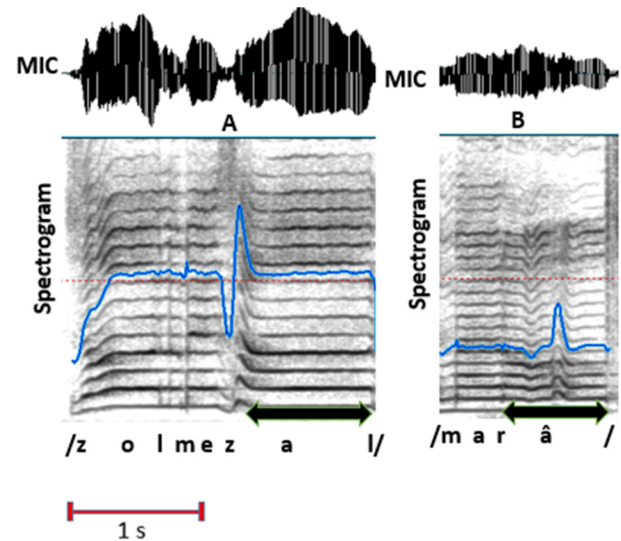

Fig.1. Pattern 1 is a single pulse tahrir pattern.

tekyeh cycles are separated in a way that a listener doesn't hear this pattern as a grouped event, but as a sequence of individual pitch fluctuations (tekyehs). Also, the singer controls the production of each tekyeh in the pattern (Fig. 2). The approximate range of the fundamental frequency extent is 3.57 ST (min of 2.60 ST, max of 6.36 ST) The pattern can have a flat baseline, a rising baseline, or a falling baseline, and the singer can stop raising or lowering the pitch and repeating cycles at any time during the production of the tahrir.

Pattern 3: This pattern is a multiple tekyeh gesture with a relatively short inter-tekyeh intervals (average 84 ms). This pattern has multiple tekyehs within a prolonged vowel and happens between two consonants and has 2 types. Type 3A is flat with no slope and has fewer tekyeh cycles in each tahrir than type 3B. Type 3B could have an ascending or a descending fundamental frequency (in the peak frequencies or the baseline frequencies from the first to the last tekyeh). As type 3A does not require a step by step increase or decrease of frequency, it is easier for the singer to produce and consequently is more frequent than type 3B (Fig. 3). The approximate range of the fundamental frequency extent is 2.95 ST, with minimum of 1.86 ST and maximum of 4.81 ST.
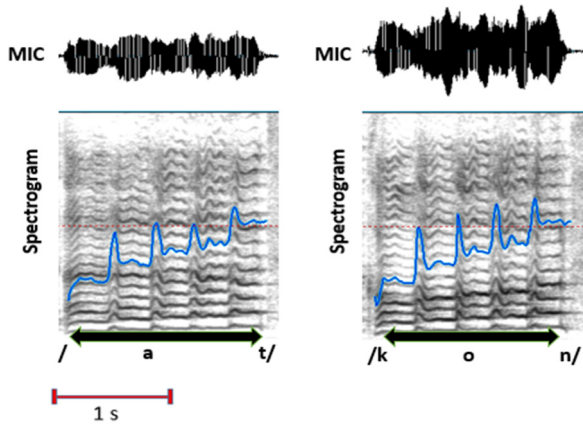
Fig.2. Pattern 2 is an ascending multiple tekyeh gesture with long plateau intervals.
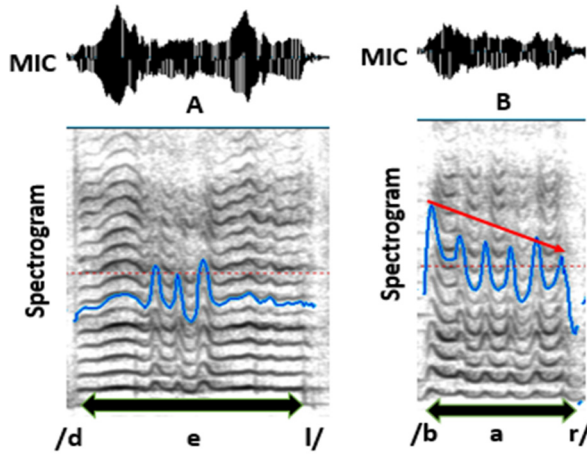

Fig.3. Pattern 3 which is a multiple tekyeh pattern could be flat (type 3A) or have slope (type 3B).

Pattern 4: This vibrato-like pattern is also a multiple tekyeh pattern with short inter-tekyeh intervals (average 44 ms) like pattern 3 but with relatively short frequency extents (fundamental frequency range of 1.57 ST, min of 0.58 ST, max of 2.15 ST), and is frequently seen during the song (Fig. 4).

Table 1. Factors contributing to categorization of tahrir patterns.

Pattern	Single/ multiple	f_o extent *	Plateau interval
1	S	Large	n/a
2	M	Large	Long
3	M	Large	Short
4	M	Small	Short

* *f_o extent means fundamental frequency increase from primary note to the secondary note in tekyeh*.
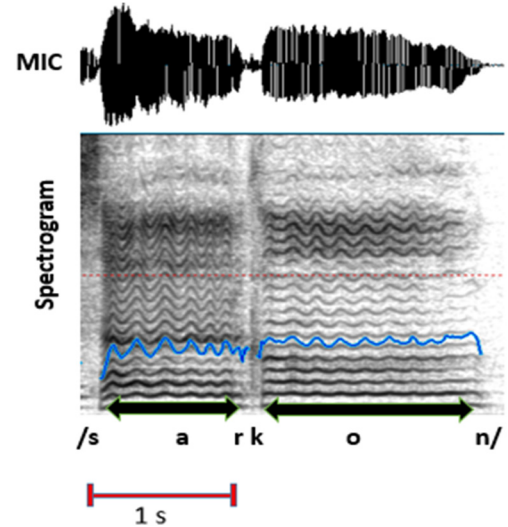

Fig.4. Pattern 4 at the end of a verse.

DISCUSSION AND CONCLUSION

For Patterns 1-3, the fundamental frequency tends to launch from a baseline pitch, rises to the peak pitch, and then lowers again to a baseline pitch, where the baseline may be the same or a different pitch, determined by the singer. The f_o dip of Pattern 1A and the vibrato-like Pattern 4 are exceptions to this observation.

Pattern 4 is similar to vibrato in western operatic singing regarding rate and extent. Rate and extent of pattern 4 are on average 8.99 Hz and 1.13 semitone, respectively and are close to the western operatic singing which has the rate of 5-7 Hz and extent of ± 1 semitone [5]. The only major difference is that the singer exerts pattern 4 intentionally in many cases like figure 4, while vibrato is not as intentional.

This study suggests that tahrir Patterns are important musically as well as functionally from a physiological and pedagogical orientations.

REFERENCES

[1]. Bahadoran, P. (2016). Analysis of Tahreer in Traditional Iranian Singing. *6th International Workshop on Folk Music Analysis*, Dublin, 15-17 June, 2016.

[2]. Castellengo, M., During, J., & Lamesh, S. (2007). The iranian tahrir: acoustical analysis of an ornamental vocal technique. *In 3rd Interdisciplinary Musicology Conference*, 31-10-2006; Paris (pp. 53-55).

[3]. Caton M. (1974). The Vocal Ornament Takiyah in Persian Music. *UCLA Selected Reports in Ethnomusicology*. 2(1).

[4]. Biglari, H. J. (2009). *Timbral and melodic characteristics of the Persian singing style of Avaz* (Doctoral dissertation in musicology, Uppsala).

[5]. Sundberg, J. (1995). Acoustic and psychoacoustic aspects of vocal vibrato. *Vibrato*, 35-62.

HARMONIC VOICE QUALITY PARAMETERS OF ROUGH AND MODAL VOWELS

Wolfgang Wokurek[1], Manfred Pützer[2]

[1] Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Germany
[2] Sprachwissenschaft und Sprachtechnologie, Universität des Saarlandes, Saarbrücken, Germany

***Keywords*: Rough and Modal Phonation Quality; Harmonic Voice Quality Parameters**

INTRODUCTION

Voice quality parameters based on amplitude measurements in the harmonic spectrum were introduced by Stevens and Hanson [1]. They notice that different aspects of phonatory quality dominate certain parts of the glottal excitation spectrum. These voice quality parameters were found to be robust under real word disturbances [2] since only little of noise power is within the narrow band of each harmonic amplitude measurement. Stevens' and Hanson's parameter definitions include a reduction of the harmonic amplitudes to compensate the resonant influence of the first formant, which might be viewed as an attempt of inverse filtering the speech sound.

Voice quality is parameterized in the present study by the bandwidth of the first formant and by spectral gradient features of the source spectrum [3, 4].

The aim of the present study is the analysis and comparison of voice quality parameters of rough and modal vowels uttered by male speakers without any known voice pathology. By using our voice quality parameters we hypothesize that a significant differentiation between vowels produced with rough and modal voice quality is possible. Furthermore, we predict that significantly different parameter values of several vowels produced with these two voice qualities can also be quoted.

METHODS

Human Data

The German carrier sentence "Der Trainer hat … gesagt" with 20 mainly monosyllabic test words was produced by 40 healthy male speakers with modal and rough voice quality. Each of the words contained at least one of the vowels [i:,e:,a:,o:,u:]. To stimulate rough voice quality speakers were asked to imagine shouting over a football pitch. Only half of the individuals finally produced enough vowels with rough voice quality for parameter analysis and comparison. Rough voice quality was auditorily approved by the examiners.

Instrumentation and Measurement

Six voice quality parameters were used [3, 4]. Five of them describe the spectral shape based on measuring amplitudes of harmonics (first, second and near the first four formants), inverse filtering, and calculating the spectral decay gradients (i.e. slopes; parameters OQGi, GOGi, SKGi, RCGi, T4Gi). In addition the relative

bandwidth of the first formant captures the damping introduced by the incomplete closed glottis (parameter IC).

RESULTS

This harmonic procedure is applicable even for rough voice quality. All voice quality parameters show significant differences between rough and modal phonation behavior (see Figure 1). Rough voice quality has a wider bandwidth of the first formant and less spectral decay.
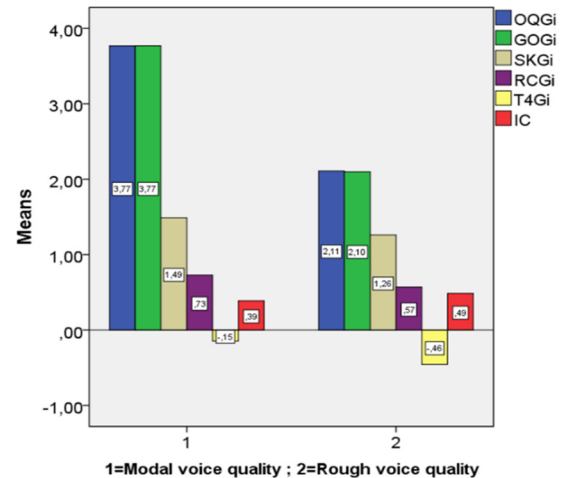

Fig. 1: Significant differentiation for the effects of condition voice quality. Parameter values (means) for modal and rough voice quality (p<0.001, Eta² .144)

Post-hoc tests (Scheffé) identify homogeneous subsets for the effects of condition vowel (5 groups). They are significantly different from each other. In the case that parameter values are listed under each subset they are not significantly different from each other. All six parameters are used as dependent variables in the statistical model.

The parameters OQGi (see Figure 2) and GOGi (see Figure 3) reveal relatively minor variations concerning the subsets and the parameter values. In Figure 2, it can be observed that front vowels [i, e] and back vowels [u, o] are part of the same subset whereas the low vowel [a] represents another subset.

Figure 3 shows the results of the post-hoc tests for the parameter GOGi. Now, four subsets appear showing a similar tendency of vowel distribution as observed for the parameter OQGi.
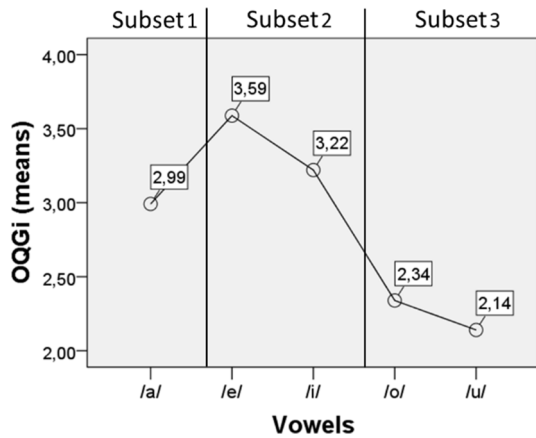
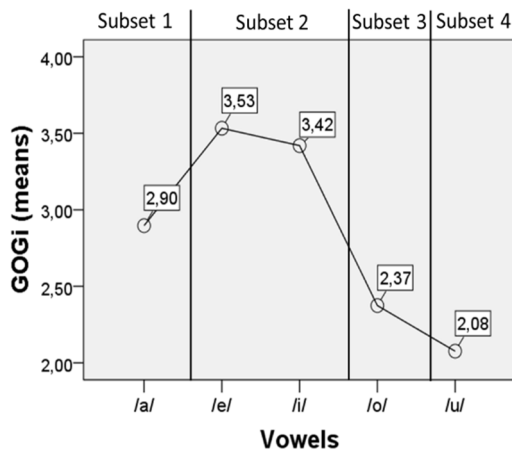**Fig. 2: Subsets of parameter values (independent variable vowel, dependent variable OQGi; p< 0.05)**

Fig. 3: Subsets of parameter values (independent variable vowel, dependent variable GQGi; p< 0.05)

Post-hoc tests for the other parameters (SKGi, RCGi, T4Gi and IC) also reveal 3 and 4 subsets, respectively. In these subsets vowel distribution patterns containing contrary places of articulation (front - back; low - high) appear not to the same extent as shown for the parameters OQGi and GOGi.

DISCUSSION AND CONCLUSION

Applying the harmonic voice quality parameter estimation to speech frames of rough and modal voice quality yielded a number of interpretable results. On the other hand a number of speech frames were excluded from the statistical analysis due to error conditions indicating insufficient harmonic structure or mismatch of the spectral decay model. Useful parameters for these frames are still worth looking for.

Considering the analysis method used in the present study the obtain statistically significant differences between rough and modal voice quality may point to the following: The lower parameter values (OQGi, GOGi, SKGi, RCGi, T4Gi) for rough phonatory behavior (see Fig.

1, right column) suggest a relative flat spectrum (spectral tilt) and less spectral decay when this spectrum is comparted to the spectrum of modal voice quality. However, the parameter values of modal voice quality are higher (see Fig. 1, left column). They refer to a steeper spectrum (spectral tilt) and more spectral decay. Physiologically, the lower parameter values of rough voice quality may reflect a tendency of *phonatory effort* constituted by increased subglottal pressure and increased tension of the vocal folds. Both may cause a more rapid closing phase, hence a flatter spectrum expressed by lower parameter values. Beside these spectral source parameters the parameter IC reflecting the bandwidth of the first formant point to more incompleteness of the closure phase of the phonatory cycle. Significantly higher values in comparison to the values for modal phonation behavior can be verified when rough voice quality was produced. Further and surprisingly, post-hoc tests reveal significantly different homogeneous subsets for the effects of condition vowel. Finally, vowel distribution patterns containing contrary places of articulation (front - back; low - high) appear, especially for the two parameters OQGi and GOGi. It seems that the analysis method used in this study allows statements about the influence of different production configurations (concerning the tongue, lips and jaw position) on laryngeal phonatory behavior. This influence can be demonstrated for the effects of condition vowel when voice quality can be significantly differentiated by dependent variables (e.g., OQGi).

The present analysis method seems to be a promising candidate for observation of changes in voice quality. The observations may closely reflect the (patho)physiological behaviour of vocal fold vibration.

REFERENCES

[1] Stevens KM, Hanson HM. Classification of glottal vibration from acoustic measurements, In: *Vocal Fold Physiology*, O. Fujimura and M. Hirano (Eds), Cambridge MA: Hiltop University Press, 1998, pp. 147–170.

[2] Lugger M, Yang B, Wokurek W. Robust estimation of voice quality parameters under real world disturbances. Proceedings IEEE International Conference on Acoustics Speech and Signal Processing, vol. 1, May 2006, pp. 1097-1100.

[3] Wokurek W, Pützer, M. Automated corpus based spectral measurement of voice quality parameters. Proceedings ICPhS XV 2003, August 3-9, pp. 2173-2176, Barcelona, Spain.

[4] Wokurek W, Pützer M. Phonation Quality Detection on the Saarbrücken Voice database using Harmonic Spectrum-based Parameters. Proceedings MAVEBA2017, December 13-15, pp. 13-18, Firenze, Italy.

CHARACTERIZATION OF ACUTE EXPOSURE TO GLUCOCORTICOID STEROIDS IN HUMAN AND RABBIT PRIMARY VOCAL FOLD EPITHELIAL CELLS

Carol Xu MD PhD[1], Emily E. Kimball MS[2], Lea Sayce DPhil[1], Bernard Rousseau MMHC PhD CCC-SLP[1-3]

Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA [1]
Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA. [2]
[3] Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA

Keywords: Dysphonia; Epithelial Barrier; Steroids; Tight Junctions

INTRODUCTION

Dysphonia, or hoarseness, is a common symptom of vocal abuse, affecting millions of Americans each year. In general practice, glucocorticoid steroids are often administered to treat acute episodes of dysphonia due to their anti-inflammatory properties; however, the tissue-specific effect of steroid treatment in the vocal folds has not been characterized. In the current study, we investigated the cellular and molecular effects of glucocorticoid exposure in primary vocal fold epithelial cells derived from healthy rabbit and human laryngeal tissue[1,2].

METHODS

Cell Culture

Primary rabbit and human vocal fold epithelial cells were isolated from healthy laryngeal tissue according to approved Vanderbilt IACUC protocols and cultured to passage 2 (P2). The cells were treated for 10 days starting with P2 day 4 with a low, medium, or high dose of three different glucocorticoid steroids: dexamethasone (0.1, 1, 10 μg/ml), methylprednisolone (1, 10, 100 μg/ml), and triamcinolone (1, 10, 100 μg/ml).

Data Collection

P2 epithelial cells were cultured on a collagen IV coated porous cell culture insert suspended in a 6-well plate. TEER was measured using an epithelial voltohmmeter. For the experiments described, TEER was measured daily from P2 Day 4 to P2 Day 10 , and cell layers were fixed on day 11. Cell culture inserts were fixed with methanol on P2 day 11, permeabilized with 0.1% Triton X 100 and EtOH/Acetic Acid, and immunolabeled with antibodies against E-cadherin, occludin, and ZO-1.

Analysis

For each glucocorticoid treatment condition, immunofluorescence of ZO-1, occludin, and E-cadherin was imaged with a Nikon 90i microscope at 20x, and analyzed using ImageJ [3]. Three images were collected per condition, and images were randomized and manually thresholded. This intensity threshold was used to calculate percent positive area of each protein of interest in each glucocorticoid condition.

RESULTS

TEER

Treatment of vocal fold epithelial cells with a high dose of methylprednisolone correlated with a decrease in TEER in both rabbit and human primary epithelial cells. The negative effect of methylprednisolone on TEER was dose dependent. Dexamethasone and triamcinolone did not yield large changes in TEER compared to the glucocorticoid-free control condition. Condition-specific changes in TEER were similar across human and rabbit vocal fold epithelial cells.
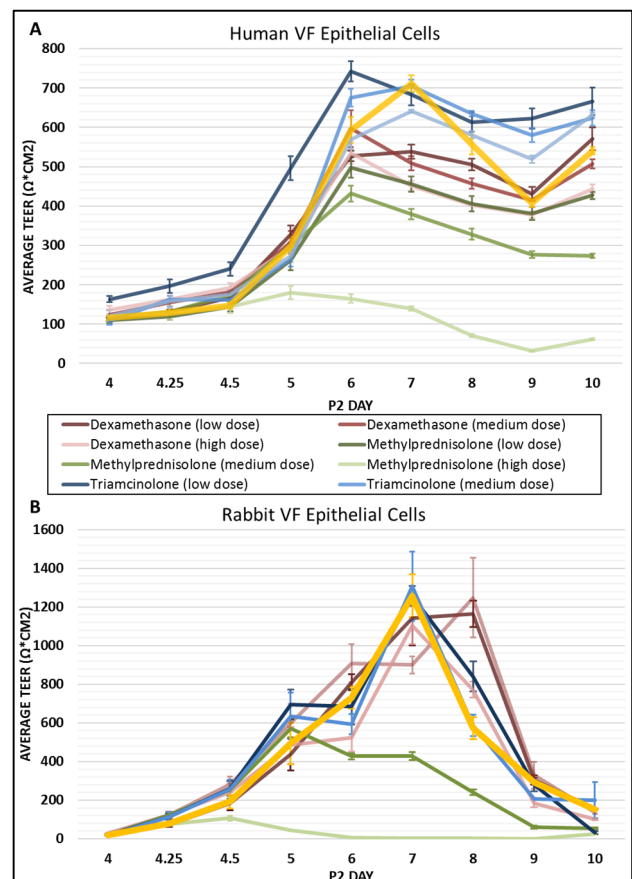

Fig. 1 Human (A) and Rabbit (B) TEER

Human Epithelial Cell Tight Junction Proteins

Human vocal fold primary epithelial cells exhibited decreased expression of ZO-1 and occludin with increasing doses of methylprednisolone treatment. Human vocal fold primary epithelial cells demonstrated increased expression of E-cadherin with decreasing doses of dexamethasone.
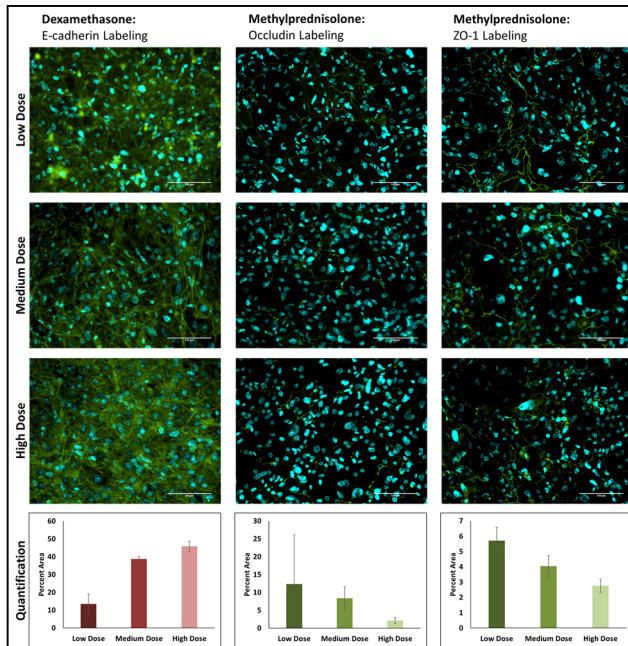

Fig. 2 Tight junctional protein immunofluorescence in human VF epithelial cells. Scale bar 100µm.

DISCUSSION

The vocal fold epithelium has emerged as an important vocal fold layer involved in critical functions, such as active ion transport, and maintenance of surface fluid composition. In many vocal fold diseases, glucocorticoids are the mainstay treatment**[4,5]**. Using the classical cell culture chamber approach, it was found that treatment of vocal fold epithelial cells with methylprednisolone correlated with a decrease in TEER in both rabbit and human primary epithelial cells**[6]**.

Transepithelial electrical resistance is mediated in part by junctional complex proteins such as ZO-1, occludin, β-catenin, and E-cadherin. Consistent with TEER results, methylprednisolone correlated with decreased ZO-1 and occludin expression. However, low levels of dexamethasone correlated with increased E-cadherin expression with no corresponding change in TEER. It is apparent from the current study that the effect of glucocorticoids on vocal fold transepithelial electrical resistance and intercellular tight junctions is a complex process that warrants further investigation.

CONCLUSION

Treatment with medium and high doses of methylprednisolone resulted in decreased TEER in both rabbit and human primary vocal fold epithelial cell culture. High doses of methylprednisolone also correlated with decreased expression of tight junction proteins ZO-1 and occludin expression in human cells. These findings suggest functional and structural changes to the vocal fold epithelial barrier, which could be advantageous for drug selection, delivery, and the treatment of acute phonotraumatic injuries.

Future studies will aim to replicate these findings in additional cell lines and to quantify the effect of glucocorticoids on apoptosis, proliferation, gene transcript levels, and cytokine production and secretion in vocal fold epithelial cell primary culture.

REFERENCES

[1] Rafii B, Sridharan S, Taliercio S, et al. Glucocorticoids in laryngology: a review. *The Laryngoscope*. 2014;124(7):1668-1673.

[2] Campagnolo AM, Tsuji DH, Sennes LU, Imamura R. Steroid injection in chronic inflammatory vocal fold disorders, literature review. *Braz J Otorhinolaryngol*. 2008;74(6):926-932.

[3] CA Schneider, WS Rasband, and KW Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 671–675, Jul. 2012. [3]

[4] Mortensen M, Woo P. Office steroid injections of the larynx. *The Laryngoscope*. 2006; 116(10):1735-1739.

[4] Tateya I, Omori K, Kojima H, Hirano S, Kaneko K, Ito J. Steroid injection to vocal nodules using fiberoptic laryngeal surgery under topical anesthesia. *Eur Arch Oto-Rhino-Laryngol Off J Eur Fed Oto-Rhino-Laryngol Soc EUFOS Affil Ger Soc Oto-Rhino-Laryngol - Head Neck Surg*. 2004;261(9):489-492.

[5] Sivasankar M, Erickson E, Rosenblatt M, Branski RC. Hypertonic Challenge to Porcine Vocal Folds: Ef-fects on Epithelial Barrier Function. *Otolaryngol -- Head Neck Surg*. 2010;142(1):79-84.

THE PERCEPTION OF EFFORT AND THE EXPERIENCE OF PHYSICAL TENSION IN FLUENCY DISORDERS

J. Scott Yaruss[1]

[1] Communicative Sciences and Disorders, Michigan State University, East Lansing, MI, USA

***Keywords*: Stuttering, Fluency Disorders, Physical Tension, Struggle**

INTRODUCTION

Individuals who stutter often report that speaking is effortful for them. Indeed, observable physical tension is one of the defining behaviors commonly associated with stuttering. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [1] includes, among other diagnostic criteria, "words pronounced with an excess of physical tension," as one of the key symptoms of stuttering. Many definitions of stuttering behavior incorporate descriptions of physical tension [2, 3], and stuttering severity indices include ratings of observable physical tension. For example, the widely used Stuttering Severity Instrument [4] is based, in part, on descriptions of "physical concomitant behaviors" that reflect effort and struggle during moments of stuttering. Moreover, numerous treatment techniques (including so-called "stuttering modification strategies) have been developed in an attempt to reduce physical tension during moments of stuttering [e.g., 5].

Examples of the physical tension and struggle behaviors commonly seen in people who stutter include tension of the facial muscles, facial grimacing, eye blinking, pursing of the lips, and tensing of the cheeks. Tension can even be observed in peripheral movements, such as clenching of the fists, tapping of the feet, and tightness in the abdomen or other parts of the body [6, 7].

Although struggle behaviors are frequently observed in those who stutter, the specific reasons for increased physical tension—during both perceptibly fluent and stuttered speech—are not clear. Specifically, it is not yet known whether physical tension is the direct result of a primary motor impairment that underlies stuttering behavior or whether the physical tension is a *reaction* to an underlying impairment, as the speaker attempts to cope with an underlying sensation of being unable to speak [7].

It is also not clear why people who stutter report that speaking is effortful, or even whether the sensation of effort is universal amongst those who stutter. Certainly, some aspects of perceived effort are related to excess physical tension in the speech musculature, but other aspects may be associated with underlying language formulation and speech production difficulties that may ultimately cause stuttering [8]—or with resulting anxiety that affects many individuals who stutter [e.g., 9, 10].

To date, limited research has been conducted on this topic, though some have sought to examine the problem of physical tension and physical effort from multiple perspectives, using both instrumental and qualitative (self-report) methods (see review in [7]).

Given that stuttering is a neurological condition that results in disrupted speech behaviors [e.g., 8], it is widely assumed that people who stutter exhibit some type of underlying impairment in their motor system. Some aspects of physical tension may be the direct result of using this impaired motor system, though the link is not obvious. Research has shown that people who stutter exhibit subtle differences in motor coordination when compared to those who do not stutter, and the perception of effort (and the resulting physical struggle) may simply result from the difficulty in coordinating movement using this impaired system. The instrumental measures used thus far have not clarified how the underlying physical impairment may be related to the production of speech disfluencies. Therefore, further research on the nature of any underlying impairment is needed.

Other aspects of physical tension appear to result from a speaker's reactions to the sensation of being "stuck," either immediately before or during the moment of stuttering [11]. In qualitative studies, people who stutter affirm their perception that the moment of stuttering actually begins before the surface manifestations become visible to observers [7, 12]. In other words, speakers sense that something has "gone wrong" somewhere in the complicated processes of language formulation and speech preparation before the disruption in speech becomes evident. This sensation (often called a "loss of control," [11] though the term is not without controversy) contributes to the sense of effort and difficulty in speaking. Speakers explain that they may tense their muscles or struggle with their speech in response to this sensation of loss of control, as they attempt to force words out [7]. Again, the nature of the loss of control and the perception of effort is not well-understood, so further research on the speaker's sensations associated with stuttering is also needed.

This presentation explores the perception of stuttering as reported by individuals who stutter and raises questions for ongoing and future research about how to improve our understanding of effort and physical tension during the moment of stuttering.

It is particularly important to recognize two basic facts that have been revealed through recent research that has sought the opinions of people who stutter about physical tension during stuttering and about the nature of the moment of stuttering itself:

First, people who stutter perceive physical tension that cannot be reliably observed by listeners, including trained speech-language pathologists. For example, people who stutter routinely report physical tension in the abdomen, even though such perceptions cannot be reliably measured by observers [6, 7]. Moreover, people who stutter also report that speaking can be effortful for them even when they are not exhibiting observable physical tension [12]. Thus, there is more to tension and struggle during stuttering than has been described in common measures of physical tension. A more comprehensive definition of stuttering is needed in order to capture these real-world experiences of people who stutter.

Second, people who stutter perceive that the moment of stuttering begins before any type of atypical behavior becomes visible to an observer. In other words, people who stutter may feel that they are stuttering even when listeners cannot identify that stuttering is taking place. Thus, definitions of stuttering that are based on observable behavior alone are not sufficient for understanding the nature of the struggle that people who stutter actually experience [11, 12].

As important as these observations about the nature of physical tension and struggle may be, it is even more important to reflect on the fact that the vast majority of research in the field of fluency disorders has been based on observations of people who stutter by listeners or conversational partners (see detailed reviews in [3]). The reason for this is understandable, given that listeners have been historically believed to be more objective judges of behavior than speakers themselves. Still, the finding that even skilled listeners are unable to identify key aspects of the experience of stuttering that are consistently reported by speakers means that future research will need to more carefully consider the reports and experiences of those who stutter. This is particularly true with respect to reports of speaking effort.

To date, however, no studies have specifically examined speaker perceptions of effort during either fluent or disfluent speech. A new initiative at Michigan State University, known as the "Spartan Stuttering Surveys" (http://surveys.StutteringCenter.org) aims to address this shortcoming by asking novel questions of people who stutter about their experiences of moment of stuttering and the broader stuttering disorder, and their perception of various types of disfluencies, physical tension and struggle during fluent and disfluent speech, and speaking effort. The results from this work will increase our understanding of the nature of stuttering and, ultimately, lead to improvements in stuttering theory and therapy.

REFERENCES

[1] American Psychiatric Association. *Diagnostic and Statistical Manual for Mental Disorders, 5th ed.* Washington, DC: American Psychiatric Association; 2013.

[2] Wingate M. A standard definition of stuttering. *J Speech Hear Disord* 1964;29:484-489.

[3] Bloodstein O, Bernstein Ratner N. *A handbook on stuttering, 6th ed.* New York: Thompson-Delmar; 2008.

[4] Riley G. *Stuttering Severity Instrument 4th ed.* Austin TX: Pro-Ed; 2009.

[5] Van Riper C. *The treatment of stuttering.* Englewood Cliffs: Prentice Hall; 1973.

[6] Snidecor J. Tension and facial appearance in stuttering. In W Johnson (ed), *Stuttering in Children and Adults.* Minnesota: University of Minnesota Press; 1955.

[7] Tichenor S, Leslie P, Shaiman S, Yaruss JS. Speaker and observer perceptions of physical tension. *Folia Phoniatrica et Logopedica* 2017; 69: 180-189.

[8] Smith A, Weber C. How stuttering develops: The multifactorial dynamic pathways theory. *J Speech Lang Hear Res* 2017;60(9):2483-2505.

[9] Iverach L. Social anxiety disorder and stuttering: Current status and future directions. *J Fluency Dis* 2014; 40:69-82.

[10] Yaruss JS, Quesal RW. Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *J Fluency Dis* 2006;31:90-115.

[11] Perkins W. What is stuttering? *J Speech Hear Dis* 1990;55:370-382.

[12] Tichenor S, Yaruss JS. A phenomenological analysis of the moment of stuttering. *Am J Speech-Lang Path* in press.

PHONATION AEROACOUSTIC SOURCE STRENGTHS

Feimi Yu[1], Gage Walters[2], Michael Krane[2], and Lucy Zhang[1]

[1] Mechanical Engineering Department, Rensselaer Polytechnic Institute, Troy, NY, USA
[2] Applied Research Laboratory, Penn State University, State College, PA, USA

***Keywords*: Acoustics; Aerodynamics and kinematics; Modeling**

INTRODUCTION

This presentation presents estimates of the principal phonatory aeroacoustic source. Glottal volume flow [1], .vocal fold drag [2], and glottal jet dynamic pressure [3] have been used to quantify this source. Krane [3] showed their equivalence, using scaling arguments, and further that the transglottal pressure (TGP) force, because it is nearly equal and opposite to vocal fold drag, can be used as a surrogate for the latter. McPhail *et al.* [4] presented supporting measurements in a physical model of phonation. This presentation provides further support for these claims. First, a high-fidelity aeroelastic-aeroacoustic simulation of phonation (see [5] for details) shows that the TGP force and the vocal fold drag are nearly equal and opposite. Measurements in the same physical model used in [4], this time using vocal fold models exhibiting qualitatively different vibration patterns, further establish the relationship between sound radiated out of the model vocal tract, and the fluctuating part of the TGP.

METHODS

High-fidelity simulation

To compute the relevant terms in the equation of motion, a high-fidelity aeroelastic-aeroacoustic simulation of phonation in an idealized geometry was performed. For further details on the simulation, see [5].

Control volume analysis

The results of the high-fidelity simulation were used to compute the terms in the integral momentum equation for the laryngeal control volume:

$$\rho_0 \frac{d}{dt} \iiint_V u_1 dV + \oiint_S \rho_0 u_1((u_1 - w_1)n_1) dS = \underbrace{(p_A - p_D)S}_{\substack{\text{3. transglottal pressure}\\\text{force}}} - \underbrace{D_{VF}}_{\substack{\text{4. vocal fold}\\\text{pressure}\\\text{drag}}} - \underbrace{F_f}_{\substack{\text{5. vocal fold}\\\text{friction drag}}} \quad (1)$$

1. rate of change of momentum 2. net momentum efflux

Physical model experiments

The measurements were performed in the Penn State upper airway model [4, 6], shown in Figure 1. The model has a uniform, 7.78 cm^2 square cross-section with vocal tract length of 17.2cm. The larynx section houses two 3D-printed plastic brackets onto which silicone rubber vocal folds are molded [7,8]. Three vocal fold models, also shown in Fig. 1, were fabricated for these experiments: (1) a two-layer "M5" shape [9,10], (2) a 2-layer "M5 ligament" model with an additional layer of paper molded between

body and cover layers to provide vertical stiffness, and (3) an "Ellipse" model inspired by MRI measurement [10], comprised of body, "ligament", and cover layers.

As shown in Fig. 1, a high-speed camera was used to image the vocal folds. Mouth pressure was measured using a flush-mounted microphone, and sub- and supraglottal pressures were measured using Kulite XCS-093 pressure transducers. Finally, a Rothenberg mask was used to measure mouth volume flow. For each subglottal pressure, 5s of data was acquired, with all instruments sampling at 22.1kHz. This was repeated over a range of subglottal pressures.

Analysis

Image analysis of the high-speed video was used to estimate waveforms of projected glottal area. TGP was computed as the difference between the outputs of the Kulite transducers. TGP mean and standard deviations were computed, as well as standard deviation of the mouth pressure.
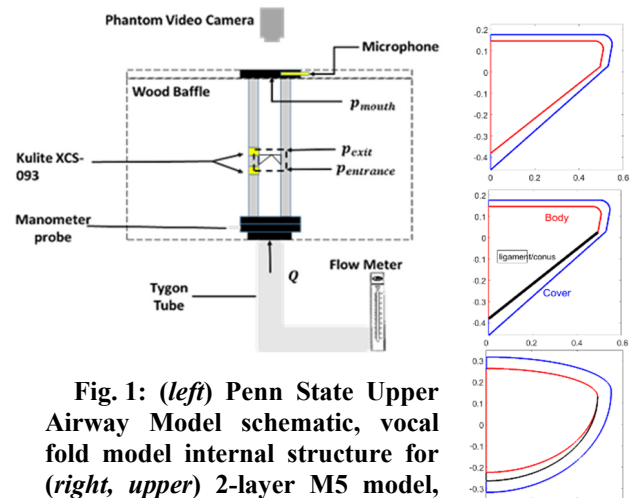

Fig. 1: (*left*) Penn State Upper Airway Model schematic, vocal fold model internal structure for (*right, upper*) 2-layer M5 model, (*right, mid*) 2-layer M5 + ligament model, (*right, lower*) elliptical model. Same models were used in [5].

RESULTS AND DISCUSSION

Estimates of the terms in Eqn. 1 from the high-fidelity simulations are shown in Fig. 2. It is clear that all other are

negligible compared to vocal fold drag (plotted here with a negative sign, for clarity) and TGP force.
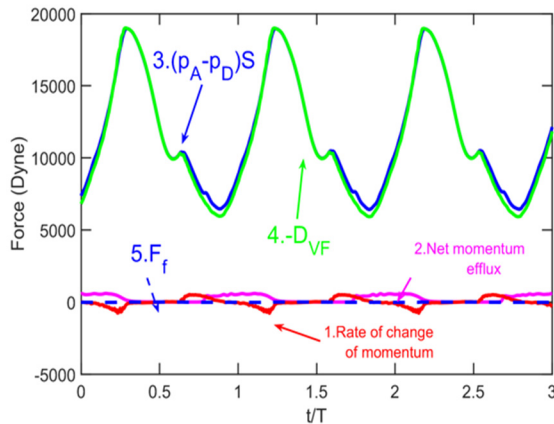

Fig.2: Waveforms of integral momentum equation terms for laryngeal control volume, from high-fidelity simulation.
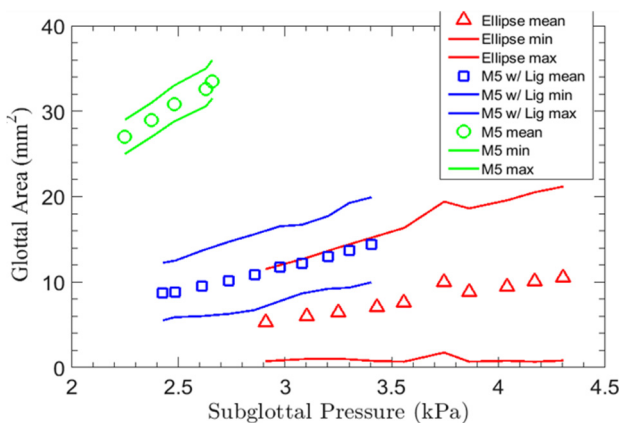

Fig.3: Glottal open area for physical models. Symbols indicate mean area, lines maximum and minimum area. Same patterns exhibited in [5].
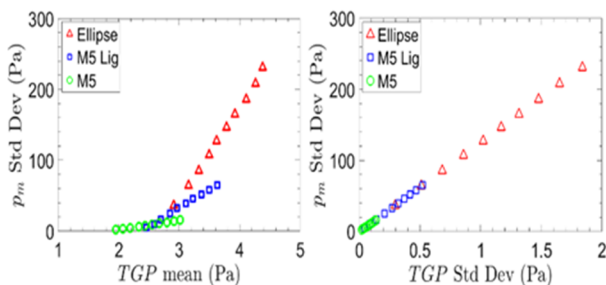

Fig.4: Aeroacoustic behavior of physical models. (*left***) mouth sound pressure p_m vs. mean *TGP*, (***right***) p_m vs. fluctuating *TGP*.**

Figure 3 shows a plot of mean glottal area vs. mean TGP, as well as maximum and minimum area, for the three vocal fold models tested. The three models clearly exhibit

different vibration patterns. Estimates of mouth sound pressure vs. mean TGP and TGP standard deviation are shown in Figure 4. Fig. 5 (*left*) shows how sound production improves as degree of closure increases. Fig. 5 (*right*) shows a strong correlation between mouth sound pressure and fluctuating TGP, over a wide range of vibration patterns, establishing TGP, as well as vocal fold drag, as measures of phonation source strength.

CONCLUSION

Vocal fold drag and TGP force are so nearly equal and opposite that they can be used interchangeably to quantify phonation aeroacoustic source strength. This was demonstrated for a range of vibration patterns. Correlation of fluctuating TGP to fluctuating glottal volume flow will further establish equivalence of the theoretical descriptions of phonation aeroacoustics.

ACKNOWLEDGMENTS

REFERENCES

[1] Fant, G. *Acoustic Theory of Speech Production* (Mouton, the Hague, 1960), Appendix A.2, 265-280

[2] Howe, M. and McGowan, R. "Sound generated by aerodynamic sources near a deformable body, with application to voiced speech," *J. Fluid Mech*. 2007; **592**, 367-392.

[3] Krane, M. Aeroacoustic sources in phonation – a control volume approach, subm. *J Acoust Soc Am*. 2018.

[4] McPhail, M., Campo, E., Krane, M. Aeroacoustic source characterization in a physical model of phonation. subm. *J Acoust Soc Am*. 2018.

[5] Krane, M., Walters, G., Yu, F., and Zhang, L. Voice energy utilization and efficiency, *Proceedings of the 11th ICVPB*, E.Lansing, MI, Aug. 2018.

[6] Campo, E. "The effect of vocal fold geometry on fluid structure acoustic interactions in an experimental model of the human airway," MS thesis, Penn State University, 2012.

[7] Riede, T., Tokuda, I., Munger, J., and Thomson, S. Mammalian laryngseal air sacs add variability to the vocal tract impedance: Physical and computational modeling," *J Acoust Soc Am* 2008 ; 124(1), 634-647.

[8] Drechsel, J. Characterization of synthetic, self-oscillating vocal fold models, MS thesis, Brigham Young University-Provo, 2008.

[9] Scherer, R., Shinwari, D., De Witt, K., Zhang, C., Kucinschi, B., and Afjeh, A. Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," *J Acoust Soc Am* 2001; 109(4), 1616-1630.

[10] Pickup, B. & Thomson, S. Flow-induced vibratory response of idealized versus magnetic resonance imaging-based synthetic vocal fold models. *J Acoust Soc Am*., 2010; 128(3), 124-129.