Classification of sEMG Signals for the Detection of Vocal Fatigue based on VFI Scores

Yixiang Gao[1]; Maria Dietrich[2]; Melinda Pfeiffer[2] and G. N. DeSouza[1]
[1]ViGIR Lab, Dept. of Electrical Engineering and Computer Science
[2]Dept. of Communication Science and Disorders
University of Missouri
Email: yg5d6@mail.missouri.edu[1], DietrichM@health.missouri.edu[2]
mlpr5b@mail.missouri.edu[2], DeSouzaG@missouri.edu[1]

***Abstract*—In this new research, we expand on our previous system for vocal fatigue detection by adding five new features in the classifier. We also perform further testing on 37 test subjects. The goals were: 1) to classify subjects performing normal versus simulated pressed vocal gestures; 2) to distinguish vocally healthy from vocally fatigued subjects as determined by VFI score on factor 1; and 3) to determine the validity of the labels vis-a-vis the choice of this same VFI-factor-1 boundary. As the results demonstrated, the choice of classifier and the new features were quite appropriate, while there is margin for better choices of the VFI-factor-1 boundary.**

I. Introduction

Voice disorders are a serious threat to a teacher's career. Teachers with a history of voice problems while student teachers have an almost nine-fold greater chance of developing a voice disorder [1, 2]. However, while symptoms of vocal effort and vocal fatigue emerge with teaching, signs of these symptoms are often elusive during clinical evaluations. The aim of this paper is to demonstrate: (i) that it is possible to use surface EMG (sEMG) of the anterior neck to classify voice productions. That is, to separate normal voice productions from voices known to have symptoms of vocal fatigue based on the Vocal Fatigue Index (VFI), a self-report questionnaire [3]; and (ii) that the proposed classification system is able to achieve preliminarily detection of clinical vocal fatigue based on the VFI.

II. Background

A. Vocal Fatigue

The most common vocal symptoms in teachers are vocal fatigue, vocal effort, and hoarseness [4]. Such symptoms are typical for hyperfunctional voice disorders, also known as muscle tension dysphonia (MTD). In the meantime, the rise of wireless and ambulatory monitoring devices for sEMG facilitates the study of extralaryngeal processes underlying vocal fatigue. The perception of increased vocal effort/fatigue and measures of extralaryngeal activity are intricately linked [5]. Vocal effort is believed to be in part the result of compensatory extralaryngeal activity to maintain adequate voicing during vocal fatigue[6]. However, the detection of vocal fatigue and early signs of MTD are often evasive [7], both in the screening and clinical setting. Consequently, there is a need to improve early detection of preclinical voice problems.

B. sEMG Pattern Recognition

In our first study presented in [8], a hierarchical classifier based on source signal separation, named HiGUSSS [9], detected vocal gestures using four sEMG channels positioned at the anterior neck. In that study, six normal voice gestures were tested and the system achieved overall accuracy of 85%. More recently in [10], the same classifier was applied to a larger set of both normal and simulated pressed voice gestures, from ten test subjects. A total of one hundred gestures were considered for this second experiment and the classifier achieved overall accuracy of 85% for classifying the ten distinct gestures and 95% for classifying between the normal and simulated pressed voice groups.

In this paper, we continue those studies by expanding on our selection of subjects to include individuals with actual vocal fatigue.

III. Experimental Setup and Proposed Method

A. Test Subjects

This study included data from 37 females, ages 21-39 years. All subjects were in good general and vocal health, by self-report. None of the subjects had a history of voice disorders or vocal fold lesions per laryngeal videostroboscopy (M.D.), based on agreement by M.D. and an otolaryngologist experienced in voice (M.P.). Each subject was asked to complete the Vocal Fatigue Index (VFI) questionnaire [3]. The questionnaire was completed on the first factor, Tiredness and Avoidance, during pre-screening and then again in full during the experiment. The experimental procedures involving human subjects described in this paper were approved by the University of Missouri Institutional Review Board. In this paper, we used the VFI scores on the first factor from the day of the experiment. Among all 37 subjects, the mean VFI score on factor 1 was 5.703 with a standard deviation of 6.046. However, we used the mean and standard deviation data from [3] for assigning the *Vocally Healthy* and *Vocally Fatigued* labels.

Figure 1. The system setup for data collection

B. Data Acquisition

Figure 1 illustrates the system setup for data collection. The equipment was installed in an audiology soundbooth (IAC Acoustics, Bronx, New York). The signals captured by a head-worn condenser microphone (Model C520, AKG, Harman, Austria) were sent to an audio amplifier (Scarlett i2i, Focusrite, High Wycombe, UK) before being sampled into digital form and saved by a 16/35 PowerLab (ADInstruments, Dunedin, New Zealand).

At the same time, four sEMG signals were collected by wireless sensors (Trigno^{TM} Mini, Delsys, USA) connected to a base station. According to the manufacturer, this type of sensor is suited for recording sEMG on small and "difficult-to-isolate" muscles. For this study, we used the same electrode placement as in [10]. Both the audio amplifier and the sEMG base station were connected to the PowerLab, which performed simultaneous sampling of all inputs. The sampling rate was set to 4kHz for the sEMG signal and 20kHz for audio signals with 16-bit quantization. In this study we focus on the sEMG data only.

C. Data Collection Protocol

After each subject completed the questionnaire (VFI) and received training by a certified speech-language pathologist (M.D.) to produce vowels with a pressed voice, we used a caliper (Lange Skinfold Caliper, Beta Technology, Cambridge, MD) to measure their skinfold thickness overlying the submental and infrahyoid muscles, where the electrodes were placed. This measurement was performed three times to obtain an average thickness value. Next, each subject started the data collection (both sEMG and acoustic) by producing a series of normal and simulated pressed voice gestures (vowels) broken up by syllables and sentences. At the end, we asked each subject to press their chin on a dynamometer (Chatillon LG-050 with curved compression fixture SPK-FMG-142, Ametek, Largo, FL) to collect sEMG data during maximum voluntary contraction (MVC) and submaximal voluntary contraction (50% MVC) with 1-min rest intervals between collections. This entire sequence is detailed in Table I.

Table I
EXPERIMENT PROTOCOL FOR sEMG VOICE DATA COLLECTION.

Task	Description	Reps	Time
baseline	Neutral with no movements for collecting pure noise	1	2 secs
syllable1	"afa afa afa ifi ifi ifi ufu ufu ufu"[11]	1	6 secs
/a/ normal	/a/ as in honest	55	2 secs
/u/ normal	/u/ as in you	55	2 secs
/i/ normal	/i/ as in feel	55	2 secs
sentence1	"The dew shimmered over my shiny blue shell again"[11]	55	4 secs
sentence2	"Only we feel you do fail in new fallen dew"[11]	55	4 secs
syllable2	"afa afa afa ifi ifi ifi ufu ufu ufu"[11]	1	6 secs
/a/ pressed	/a/ with a breath hold	55	2 secs
throat clear	Single throat clear	55	1.5 secs
/u/ pressed	/u/ with breath hold	55	2 secs
cough	Single cough	55	1.5 secs
/i/ pressed	/i/ with breath hold	55	2 secs
syllable3	"afa afa afa ifi ifi ifi ufu ufu ufu"[11]	1	6 secs
MVC	Maximum voluntary contraction	3	8 secs
50% MVC	Submaximal voluntary contraction	3	15 secs

D. Classification

In this study, we expended the feature vector used in our previous work [10]. We are now using: *GUSSS ratio* (GR) [12], *Mean Absolute Value* (MAV), *Zero Crossings* (ZC), *Slope Sign Changes* (SSC), *Waveform Length* (WL) [13], *Willision Amplitude* (WA) [14], *Root Mean Square* (RMS) [15] and *Auto Regressive* (AR) [16] as the feature vectors for classification. Finally, a Linear Discriminant Analysis (LDA) classifier was chosen because: 1) it does not require iterative training; and 2) it avoids potential under- or over-training [17].

IV. EXPERIMENTS AND RESULTS

Two experiments were performed: *Exp-1*, where all test subjects were asked to produce normal and simulated pressed vowels (/a/, /u/, /i/ in Table I); and *Exp-2*, where the test subjects were separated into two groups based on their VFI scores (vocally healthy and vocally fatigued groups). In Exp-1, the data were labeled *Negative* and *Positive* for the presence of simulated pressed vowels (simulated fatigue); while for Exp-2, the same labels *Negative* and *Positive* represented the actual presence of fatigue. In Exp-2, only the normal vowels were used – the number of repetitions for each gesture (vowel) is shown in the third column of Table I. All experiments were performed using ten-fold cross-validation: i.e. the average of ten runs using 90% of the data for training and 10% for testing.

A. Exp-1: Normal Vowels vs. Simulated Pressed Vowels

The goal of this experiment was to classify simulated pressed vowels as an indicator of vocal fatigue. So, two tests were performed: intra-subject and inter-subject. For the intra-subject test, due to performance errors, the number of samples in each class varied from 108-165 for Positive and

Table II
CONFUSION MATRIX FOR POSITIVE/NEGATIVE DETECTION OF SIMULATED PRESSED VOWELS USING 37 TEST SUBJECTS UNDER **INTRA-SUBJECT** APPROACH

	Actual Positive	Actual Negative
Predicted Positive	94.6%	3.3%
Predicted Negative	5.4%	96.7%
Accuracy	95.83%	

Table III
CONFUSION MATRIX FOR POSITIVE/NEGATIVE DETECTION OF SIMULATED PRESSED VOWELS USING 37 TEST SUBJECTS UNDER **INTER-SUBJECT** APPROACH

	Actual Positive	Actual Negative
Predicted Positive	62.9%	27.1%
Predicted Negative	37.1%	72.9%
Accuracy	68%	

55-165 for Negative detection of simulated pressed vowels. The results are shown in Table II.

For the inter-subject test, the total numbers of voice samples in each class were: 6,033 and 5,854 in the Positive and Negative classes, respectively. As the result in Table III shows, the overall accuracy dropped compared to the intra-subject test. This drop can be explained by the fact that among all test subjects, some individuals were actually considered to have vocal fatigue according to their VFI score on factor 1 [3]. Thus, it is reasonable to assume that the classifier could not separate the normal voice from subjects with vocal fatigue and the simulated pressed voice from vocally healthy subjects, which led to the next experiment.

B. *Exp-2: Vocally Healthy vs. Vocally Fatigued Subjects*

As indicated earlier, VFI scores on factor 1 were used as the indicator of vocal fatigue. Also, the results in [3] determined that the mean (standard deviation) of VFI scores are 24.47 (9.76) for patients with voice disorders and 5.16 (4.58) otherwise. So, for this study, we set the value VFI $\geqslant 15$ ($\approx 24.47 - 9.76$), as the boundary for *Vocally Fatigued* subjects. Similarly, a value VFI $\leqslant 10$ ($\approx 5.16 + 4.58$) was set for *Vocally Healthy* subjects. Lastly, subjects whose VFI fell between the two boundaries were put in the *Intermediate* group. The goal was to classify Vocally Healthy (Negative) and Vocally Fatigued (Positive) subjects. Accordingly, the number of subjects in the Vocally Fatigued group was 4, while the number of Vocally Healthy was 31. Therefore, the total number of samples (vowel gestures) labeled Positive was 657 and the number of samples labeled Negative was 5,046.

The result for this classification is shown in Table IV. The overall classification accuracy was 93.9%, which is much higher than the accuracy achieved in Exp-1-inter-subject. Also, unlike in Exp-1-inter-subject, the error now is less evenly distributed: i.e. 37.1% and 27.1% in Table III versus 41.9% and 1.4% in Table IV. However, both the high accuracy and the uneven error could be indicative of the

Table IV
CONFUSION MATRIX FOR POSITIVE/NEGATIVE DETECTION OF VOCALLY FATIGUED SUBJECTS AMONG 35 TEST SUBJECTS

	Actual Positive	Actual Negative
Predicted Positive	58.1%	1.4%
Predicted Negative	41.9%	98.6%
Accuracy	93.9%	
Sensitivity	0.581	
Specificity	0.986	

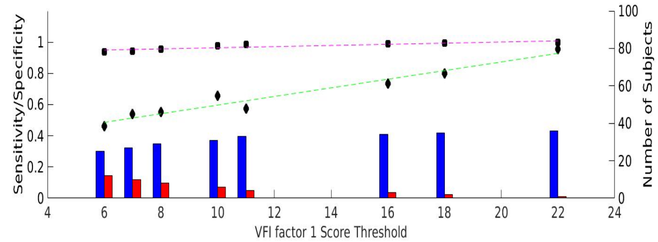

Figure 2. The sensitivity/specificity of the classification results in relation to the VFI factor 1 score threshold for splitting between positive and negative classes. The red bars indicate the number of subjects that belong to positive class and blue corresponds to negative class.

heavy bias in the dataset towards Vocally Healthy subjects: i.e. the classifier had only 657 samples to learn from vocally fatigued subjects out of 5,703 total samples. So, to better understand this result, it is worth pointing out the sensitivity (a.k.a. true positive rate) in Table IV: 0.58.

Another explanation for the low sensitivity could be in the labeling. Perhaps, the boundaries set for vocally healthy and vocally fatigued are not appropriate and subjects labeled fatigued may not be so (and vice-versa). Also, by changing these boundaries, the intermediate group – subjects that could not be assigned to either class – could provide more information for the vocally fatigued group.

To further investigate the labeling step, we set a single and variable VFI-on-factor-1 boundary to separate the vocally healthy and vocally fatigued groups. This single VFI-on-factor-1 boundary was shifted from a low limit of 5, to a high limit of 24 in incremental steps. For each step, the same classification was performed and the sensitivity was calculated.

The results for this test are depicted in Figure 2, which shows that, as the VFI boundary increases, fewer subjects are assigned to the vocally fatigued group, while the sensitivity in the classification of the vowel gestures also increases. Yet, all specificity values remain around 90%. This shows that as the VFI-on-factor-1 boundary increases, the vocal fatigue pattern becomes more distinguishable, resulting in a higher sensitivity.

V. CONCLUSION AND FUTURE WORK

In this paper, we presented our ongoing research on vocal fatigue detection. While not all outcomes are ideal, the result of 95.8% in Exp-1 indicates that the current protocol for selection of subjects, training the subjects on simulated pressed vocal gestures, feature select, and type of classifier were quite appropriate. Also, despite the small number of test subjects with vocal fatigue as indicated by their VFI scores, the use of a variable VFI boundary can be very useful in the future to better understand the dataset and, in the process, the actual labels of those same test subjects. Another interesting approach would be to automatically discover the best boundary given a set of subjects – as for example, by clustering the dataset into the two desired classes.

However, future experiments will require many more test subjects, which we are already in the process of obtaining. Once that is done, we can focus on the generalization ability of our method so as to classify hitherto unknown subjects.

Acknowledgements

The authors would like to thank: Chi-Ren Shyu, Matthew Page, Ashton Bernskoetter, Taylor Hall, Haley McCabe, Allison Walker, for their contribution to this reasearch.

The research reported in this paper was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award Number R15DC015335 to M.D.

References

[1] P. G. C. Kooijman, F. I. C. R. S. de Jong, G. Thomas, W. Huinck, K. Graamans, and H. K. Schutte, "Risk factors for voice problems in teachers," *Folia Phoniatrica et Logopaedica*, vol. 58, pp. 159–174, 2006.

[2] G. Thomas, P. G. C. Kooijman, C. W. R. J. Cremers, and F. I. C. R. S. de Jong, "A comparative study of voice complaints and risk factors for voice complaints in female student teachers and practicing teachers early in their career," *European Archives of Oto-Rhino-Laryngology*, vol. 263, pp. 370–380, 2006.

[3] C. Nanjundeswaran, B. H. Jacobson, J. Gartner-Schmidt, and K. Verdolini Abbott, "Vocal Fatigue Index (VFI): Development and validation," *Journal of Voice*, vol. 29, pp. 433 – 440, 2015.

[4] N. Roy, R. M. Merrill, S. Thibeault, S. D. Gray, and E. M. Smith, "Voice disorders in teachers and the general population: Effects on work performance, attendance, and future career choices," *Journal of Speech, Language, and Hearing Research*, vol. 47, pp. 1092–4388, 2004.

[5] N. P. Solomon, "Vocal fatigue and its relation to vocal hyperfunction," *International Journal of Speech-Language Pathology*, vol. 10, pp. 254–266, 2008.

[6] J. Spielman, E. Hunter, A. Halpern, and I. Titze. (2012) Measuring improvement in teachers with voice complaints using the inability to produce soft voice (IPSV): preliminary data. National Center for Voice and Speech Online Technical Memo 13. [Online]. Available: http://www.ncvs.org/research/NCVSOnlineTechnicalMemo13.pdf

[7] R. E. Hillman, J. H. Van Stan, D. D. Mehta, M. Zañartu, M. Ghassemi, H. A. Cheyne II, and J. V. Guttag, "Future directions in the development of ambulatory monitoring for clinical voice assessment," in *Proceedings of the 10th International Conference on Advances in Quantitative Laryngology, Voice and Speech Research*, 2013, pp. 23–24.

[8] N. R. Smith, T. Klongtruagrok, G. N. DeSouza, C. R. Shyu, M. Dietrich, and M. P. Page, "Non-invasive ambulatory monitoring of complex semg patterns and its potential application in the detection of vocal dysfunctions," in *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2014, pp. 447–452.

[9] L. A. Rivera, N. R. Smith, and G. N. DeSouza, "High-accuracy recognition of muscle activation patterns using a hierarchical classifier," in *5th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, 2014, pp. 579–584.

[10] N. R. Smith, L. A. Rivera, M. Dietrich, C. R. Shyu, M. P. Page, and G. N. DeSouza, "Detection of simulated vocal dysfunctions using complex sEMG patterns," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, pp. 787–801, 2016.

[11] Y.-A. Lien, C. Gattuccio, and C. Stepp, "Effects of phonetic context on relative fundamental frequency," *Journal of Speech, Language, and Hearing Research*, vol. 57, pp. 1259–1267, 2014.

[12] L. Rivera and G. N. DeSouza, "Recognizing hand movements from a single sEMG sensor using guided under-determined source signal separation," in *12th IEEE International Conference on Rehabilitation Robotics*, Jun 2011, ETH Zurich, Switzerland.

[13] A. J. Young, L. H. Smith, E. J. Rouse, and L. J. Hargrove, "Classification of simultaneous movements using surface EMG pattern recognition," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1250–1258, 2013.

[14] M. Zardoshti-Kermani, B. C. Wheeler, K. Badie, and R. M. Hashemi, "EMG feature evaluation for movement control of upper extremity prostheses," *IEEE Transactions on Rehabilitation Engineering*, vol. 3, pp. 324–333, 1995.

[15] T. Fukuda, J. Echeimberg, J. Pompeu, P. Lucareli, S. Garbelotti Junior, R. Gimenes, and A. Apolinário, "Root mean square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female," *Journal of Applied Research*, vol. 10, pp. 32–39, 2010.

[16] L. Hargrove, K. Englehart, and B. Hudgins, "A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control," *Biomedical Signal Processing and Control*, vol. 3, pp. 175 – 180, 2008.

[17] A. D. C. Chan and G. C. Green, "Myoelectric control development toolbox," in *30th Conference of the Canadian Medical & Biological Engineering Society*, 2007.