MONITORING VOICE CONDITION USING SMARTPHONES

F. Schaeffler, J. Beck

CASL Research Centre, Queen Margaret University Edinburgh, Scotland, UK
fschaeffler@qmu.ac.uk, jbeck@qmu.ac.uk

Abstract: **Smartphone mediated voice monitoring has the potential to support voice care by facilitating data collection, analysis and biofeedback.**

To field-test this approach we have developed a smartphone app that allows recording of voice samples alongside voice self-report data. Our long-term aim is convenient and accessible voice monitoring to prevent voice problems and disorders. Our current study focussed on the automatic detection of voice changes in healthy voices that result from common transient illnesses like colds.

We have recorded a database of approximately 700 voice samples from 62 speakers and selected a subset of 225 voice samples from 8 speakers who had submitted at least 10 recordings and reported at least one instance of a moderate cold. We extracted 12 acoustic parameters and applied multivariate statistical process control procedures (Hotelling's T^2) to detect whether instances of cold caused violations of distributional control limits.

Results showed significant association between control limit violations and reporting of a cold. While there is scope for further improvement of sensitivity and specificity of the procedure, it could already support early detection of voice problems, especially if mediated by voice experts.

Keywords: **voice problems, monitoring, acoustic analysis, smartphones**

I. INTRODUCTION

Modern smartphones offer entirely new approaches to personal health by facilitating data collection, analysis and biofeedback. This offers new methods for tackling occupational voice problems, which are endemic in some professions [1], [2].

Most occupational voice problems are behavioural (i.e. arising from ineffective voice use) [3], so can potentially be prevented through early recognition and behavioural changes. We aim to develop an early warning system for voice problems via a smartphone app, whereby people in vocally demanding professions can routinely monitor their voice and receive tailored advice if necessary. Smartphones are widely used nowadays and a number of studies suggest smartphone audio recordings can reliably be used to extract acoustic voice parameters (see e.g. [4], [5]).

Health monitoring systems often consider patterns of deviation from baseline performance as well as static thresholds. Many human physiological factors (e.g. blood pressure, body temperature) show fluctuation patterns that can be indicative of health state [6]. For voice, too, fluctuation patterns in acoustic parameters could be indicative of vocal health. To study acoustic voice fluctuation patterns we are currently recording a longitudinal database of typical and 'at risk' voices, sampled frequently over several weeks through a smartphone app. This app records voice samples and a number of voice-related self-reports alongside each recording.

To monitor voice condition in individuals we are employing statistical process and quality control procedures [7]. These procedures are designed to detect variations in patterns that indicate non-random or 'special' causes and can be applied to univariate and multivariate situations.

We assume that stability over time is an indicator of system integrity for healthy voices. Our current aim is to analyse whether acoustic parameters derived from mobile phone recordings are a) robust enough to remain stable under normal conditions, i.e. do not exceed limits expected due to normal cause variation and b) sensitive enough to pick up minor variations in the acoustic voice profile of voice users, i.e. successfully detect special cause variation that is due to changes in the user's voice.

As a test case for detecting deviations from regular voice patterns we chose instances of self-reported common colds and similar illnesses by participants, as we have so far mainly recorded speakers who do not report regular problems with their voices. Upper respiratory tract infections (URTIs), especially when accompanied by acute laryngitis, have effects on the voice that may be similar to those encountered in occupational voice problems (e.g. hoarseness, weak voice or voice loss). Successful detection of cold-related voice symptoms would therefore indicate a level of sensitivity that could support a broader range of applications.

Detection of such changes could also have more direct benefits. URTIs are a recognized risk factor in development of voice disorder [8], so if detection of cold-related changes could trigger provision of appropriate advice when most needed (e.g. reduction of voice use and techniques for reducing vocal fold impact), this could help to prevent occupational voice problems. In addition, the ability to track whether voices return to baseline after a cold may help to differentiate transient voice changes from longer lasting or chronic ones.

II. METHODS

To collect frequent voice samples from a range of speakers we developed a mobile phone app, the 'voicecheck' app, which is available for Apple iOS and Google Android devices in UK app stores. The app records audio data in uncompressed wav (pcm) format with a sampling frequency of 44 kHz, and prompts a survey alongside each recording.

For the current study the app prompted the recording of two sustained [a] vowels at a comfortable pitch and loudness, with each vowel sustained for at least 3 seconds. Afterwards participants read 9 sentences and a short passage of text (a modified and shortened version of the 'dog and duck story' [9]).

Participants were instructed to control microphone distance by holding the phone approximately a handspan (20 cm/8 inches) from their mouth.

The survey consisted of 12 questions that addressed voice use prior to recording, psychological stress, room size/configuration, current state of the participant's voice, recent throat sensations and whether the participant currently had a cold on a scale with four levels: no cold, mild, moderate, severe. For further analysis in the present study, the 'cold' variable was transformed into a binary variable by counting "no cold" or "mild cold" as 0 and all other instances of "cold" as 1.

After audio recording and survey completion, all data was securely transferred to a central server.

Participants signed up and provided consent for the project through a website (voicecheck.org.uk). After sign-up, participants received an electronic schedule of 50 recordings as a calendar (ics) file for integration into their smartphone calendar app of choice.

Success for triggering automatic reminders by this method was variable as some calendar apps did not recognise the trigger. Recording events were distributed over twelve weeks, with more intensive and less intensive weeks and 2-3 recordings per recording day. Triggers prompted recordings at 7am, 1pm and 7pm on weekdays and 9 am and 7 pm on weekends. Over the course of the project it turned out that many participants found it difficult to stick to the schedule and were therefore instructed to provide recordings whenever suitable, but leaving at least 4h between recordings.

The database currently contains around 700 recordings from 62 speakers. For the present study we selected data from 8 speakers who had completed at least 10 recordings and had recorded instances of a cold or similar illness once or more at moderate level over the course of their recordings. Table 1 provides general information about the individual speakers.

We extracted 12 acoustic parameters from the connected speech samples, using Praat [10]. Audio processing was performed in two steps, using two different Praat scripts. The first script separated sustained vowels from both sentences and connected speech, and removed pauses and unvoiced stretches from the signal, applying the method described and using parts of the script published in [11]. Only these pre-processed connected speech samples (i.e. sentences and passage of text combined) were used for further analysis in the current study.

The second script extracted the 12 acoustic parameters from the pre-processed audio files. These comprised all AVQI parameters as described in [12], using the implementation in [11]. These were smoothed cepstral peak prominence (CPPS), harmonics-to-noise ratio (HNR) as implemented in Praat, shimmer local (Shim) and shimmer local dB (ShdB), the general slope of the spectrum (Slope) and the tilt of the regression line through the spectrum (Tilt). To this we added mean F0 (Praat's cross-correlation algorithm), jitter (RAP), jitter (PPQ5), Glottal Noise Excitation Ratio [13], [14] and uncorrected (H1-H2) and corrected (H1*-H2*) first and second harmonic difference in our own implementation, following the procedure described in [15].

Prior to analysis we calculated correlations for all extracted parameters and inspected correlations of Pearson's r above 0.7. This led to the exclusion of both jitter measures as they showed high correlation with CPPS. Shim correlated highly with ShdB and the latter was kept as it showed less correlation with CPPS. H1-H2 showed high correlation with H1*-H2*. We kept the corrected version as it should provide a better estimate of harmonic energy at the glottis.

For the remaining 8 parameters we constructed multivariate Hotelling T^2 control charts using the 'hm' method and alpha-levels of .05 and .01 [7] and recorded speaker-specific upper control limit (UCL) violations. T^2 UCL violations were then compared to the presence or absence of a cold in order to see whether instances of colds and similar illnesses would affect the acoustic profile of individuals.

Performance of the procedure was evaluated by analyzing sensitivity and specificity at group and individual level.

Table 1: Speaker age range (Age), gender (Gen), smartphone type (Phone), number of recordings (Rec) and instances of cold (Cold).

Nr	Age	Gen	Phone	Rec	Cold
1	25-29	M	Samsung Galaxy S6 Edge+	33	2
2	60-64	F	iPhone 5s	66	11
3	45-49	M	iPhone 6s	34	4
4	45-49	M	HTC One (M8) & Samsung Galaxy S7 Edge	22	3
5	25-29	F	Galaxy S6	21	3
6	35-39	F	iPhone 5c & iPhone 6s	24	2
7	35-39	F	HTC One	15	2
8	25-29	M	iPhone 6	10	2
	Sum			225	29

III. RESULTS

Table 2 shows the contingency tables for presence of a cold and T^2 UCL violations across all speakers for p-values of .05 and .01. Fisher's exact test showed a significant association between cold state and UCL violations for p=.05 (p=.007) and p=.01 (p=.001). Hit rate/sensitivity for p=.05 was 62%, specificity 66%, for p=.01 sensitivity was 55%, specificity 76%.

Results for individuals show large differences in performance of the procedure. Table 3 shows individual values for sensitivity and specificity. We investigated whether individual sensitivity and specificity values were connected to the number of recordings per participant. Figure 1 shows both sensitivity and specificity as a function of the number of submitted recordings. The graph shows that specificity increases with sample size, suggesting that false alarms become rarer when speakers provide more data. Acceptable specificity values are reached for both approaches (p=.01 and p=.05) with a sample size around 30.

Table 2: Contingency table for 'hm' method and p-levels of .05 and .01

	No cold		*Cold*		*Sum*	
	.05	.01	.05	.01	.05	.01
Below UCL	129	149	11	13	140	162
Above UCL	67	47	18	16	85	63
Sum	196		29		225	

Table 3: Sensitivity and specificity per speaker for each p-level

Speaker	*Sensitivity*		*Specificity*	
	.05	*.01*	*.05*	*.01*
2	0.5	0.5	0.8	0.8
6	0.6	0.5	0.9	0.9
9	0.3	0.0	0.9	0.9
18	1.0	1.0	0.5	0.5
40	1.0	1.0	0.5	0.5
43	1.0	1.0	0.9	0.9
61	0.0	0.0	0.5	0.5
67	0.5	0.5	0.1	0.1

The pattern for sensitivity does not show a clear relationship with sample size but there is a tendency for the p=.05 method outperforming the p=.01 method with higher sample sizes.

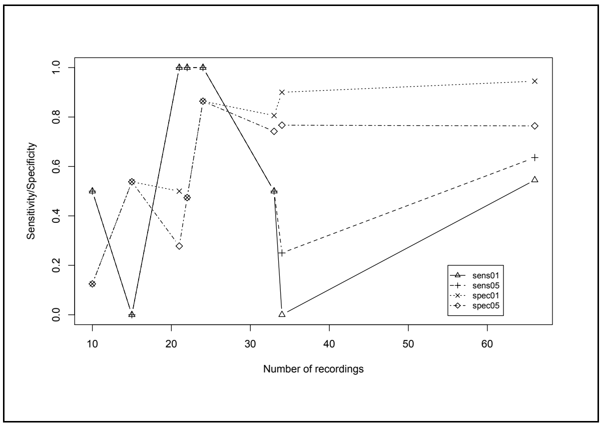

Figure 1: Changes in sensitivity and specificity of cold detection with number of recordings (sens01 – sensitivity with alpha level .01 etc).

IV. DISCUSSION

Our first analyses indicate that longitudinal monitoring of voice recordings via smartphones has potential for providing important information about the state of a voice. The current setup still generates too many misses and false alarms for unsupervised monitoring, but could be useful for supervised monitoring with voice expert support.

We have so far not excluded any recordings based on background noise levels, and we have not yet considered field effects like background noise and room size. Incorporation of these variables is likely to decrease false alarm rates in the future.

Another important future aim will be increasing the sensitivity of the method. The current acoustic parameters have not yet been analysed for their individual contributions to outlier patterns, and exclusion or addition of parameters, alongside alternative analytical approaches (e.g. machine learning) could improve hit rate.

Besides further development of the database and incorporating speakers with frequent voice problems, future research will focus on increased calibration of the method, e.g. by developing normative thresholds for acoustic parameters collected with various types of smartphones and quantifying the effects of various potential confounds that can occur in the field, e.g. background noise and room size.

V. CONCLUSION

This study presented evidence that semi-regular monitoring of voices with smartphones has potential to provide important cues about the health state of a voice. This information could be used to trigger tailored advice provided by voice experts via remote channels and thus make an important contribution to the prevention of voice problems and disorders.

VI. ACKNOWLEDGEMENTS

This study was supported by a Research Incentive Grant from the Carnegie Trust for the Universities of Scotland (grant reference 70230). We would like to thank Matthias Eichner for app programming, Tess Whittaker for help with app testing and study design and all our participants for their efforts.

VII. REFERENCES

[1] R. Martins, E. Pereira, C. Hidalgo, and E. Tavares, “Voice Disorders in Teachers. A Review,” *Journal of Voice*, vol. 28, no. 6, pp. 716–724, 2014.

[2] L. Lehto, P. Alku, T. Bäckström, and E. Vilkman, “Voice symptoms of call-centre customer service advisers experienced during a work-day and effects of a short vocal training course,” *Logopedics Phoniatrics Vocology*, vol. 30, no. 1, pp. 14–27, 2009.

[3] L. Mathieson, *The Voice and Its Disorders*, vol. 6. London: Whurr Publishers, 2001.

[4] E. Lin, J. Hornibrook, and T. Ormond, “Evaluating iPhone recordings for acoustic voice assessment,” *Folia phoniatrica et logopaedica*, vol. 64, no. 3, pp. 122–130, 2012.

[5] C. Manfredi, J. Lebacq, G. Cantarella, and J. Schoentgen, “Smartphones offer new opportunities in clinical voice research,” *Journal of Voice*, vol. 31, no. 1, pp. 111–e1, 2017.

[6] E. O'Brien, A. Coats, P. Owens, and J. Petrie, “Use and interpretation of ambulatory blood pressure monitoring: recommendations of the British Hypertension Society,” *BMJ: British Medical Journal*, vol. 320, no. 7242, p. 1128, 2000.

[7] E. Santos-Fernández, *Multivariate statistical quality control using R*. Springer Science & Business Media, 2012.

[8] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, “Voice disorders in the general population: Prevalence, risk factors, and occupational impact,” *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, 2005.

[9] A. Brown and G. J. Docherty, “Phonetic variation in dysarthric speech as a function of sampling task,” *European Journal of Disorders of Communication*, vol. 30, no. 1, pp. 17–35, 1995.

[10] P. Boersma and D. Weenink, “Praat: doing phonetics by computer (version 6.0.21) [computer software].” 2016/09/25-2016.

[11] Y. Maryn and D. Weenink, “Objective Dysphonia Measures in the Program Praat: Smoothed Cepstral Peak Prominence and Acoustic Voice Quality Index,” *Journal of Voice*, vol. 29, no. 1, pp. 35–43, 2015.

[12] Y. Maryn, M. Bodt, and N. Roy, “The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders,” *Journal of Communication Disorders*, vol. 43, no. 3, pp. 161–174, 2010.

[13] D. Michaelis, T. Gramss, and H. W. Strube, “Glottal-to-noise excitation ratio - a new measure for describing pathological voices,” *Acustica*, vol. 83, no. 4, pp. 700–706, Jul. 1997.

[14] J. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, P. Gómez-Vilda, M. Blanco-Velasco, and F. Cruz-Roldán, “The Effectiveness of the Glottal to Noise Excitation Ratio for the Screening of Voice Disorders,” *Journal of Voice*, vol. 24, no. 1, pp. 47–56, 2010.

[15] M. Iseli and A. Alwan, “An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation,” presented at the ICASSP 04, 2004, vol. 1, pp. 666–669.