Open Voice Data Project

Lizzy Alonzi, Corbin Charpentier, Travis Cline, Patrick Flor (Co-authors)
Department of Electrical Engineering and Computer Science
University of Kansas
Lawrence, KS, USA
{ealonzi,corbin1,cline,pflor}@ku.edu

Final Version

***Abstract*—Current signal and speech processing technologies allow for interesting research into and accurate diagnosis of various dysphonias. However, there is no large-scale open and accessible repository of voice samples from dysphonic patients to support such work. In this paper we discuss our work on the Open Voice Data Project (OVDP), which promotes and hosts voice data collection projects, providing interested organizations with free non-commercial access to anonymized voice data sets. We hope the OVDP will aid in elimination of data islands currently present in the field, supporting research, diagnosis, and perhaps even therapy of dysphonias. The project's initial collection project will focus on Parkinson's Disorder (PD).**

***Index Terms*—data acquisition; speech processing; mobile communication; web services;**

I. Introduction

The application of computer technologies to speech disorder research and diagnosis is an area of active research in computer science and medicine. The proliferation of mobile phones with ample compute power and advances in speech processing and machine learning give researchers the ability to collect, process, and classify voice samples at a much greater scale than ever before, leading to novel "big data" approaches to the field. Professor Max Little and his team the MIT Media Lab have presented a TED talk on his lab's approach to this burgeoning field, and the Michael J. Fox Foundation recently sponsored a Kaggle challenge (a machine learning algorithm competition with large cash prizes) on PD voice diagnosis.

Even such high-profile efforts as these, however, are hampered by a lack of one or more characteristics key for widespread usage: scale, openness, and consistency. In *Automatic Speech Signal Analysis for Clinical Diagnosis and Assessment of Speech Disorders*, Ladan Baghai-Ravary highlights "the current absence of a realistic database of remotely collected speech samples," and emphasizes that "adherence to standardised methods and datasets is shown to be crucial to the evaluation of new algorithms [for detection and classification of dysphonias]". Examining one current data set in detail, she goes on to note that "typically only 175 disordered speakers can be included out of more than 650 in the complete MEEI [Massachusetts Ear and Eye Infirmary] database" for studies addressing particular dysphonias [1].

We believe there is a better way: the technology exists to create and support access to a large-scale, open database of voice samples that would allow researchers and clinicians, even those without large grant funding, to better understand dysphonias. Such a database would link information from diverse classification systems like the Dysphonia Severity Index and Unified Parkinson's Disease Rating Scale to the waveforms themselves, enabling cross-silo research and providing an opportunity to select amply sized data sets based various combinations of specific tags and characteristics.

In this paper we provide a narrative and self-evaluation of our effort at implementing a solution to these problems, the Open Voice Data Project. OVDP will collect anonymized voice data from patients and public volunteers via a mobile application, and will provide the research community with free and open access to it via a well-structured API. Our hope is that by providing meaningfully structured, open accessibility to data on dysphonias, we can facilitate medical research and support therapist-patient interactions, ultimately leading to better understanding, diagnosis, and treatment of these conditions.

II. Related Work

As we mentioned in the Introduction, there has been much research into computer processing and classification of dysphonic voice samples. One significant player, already mentioned above, is Max Little and his research lab [2]–[5]. Others are research groups at Trier University and Erlangen University Clinic in Germany [6], [7] We have focused on the mobile/web *collection* and open hosting of dysphonic voice data on a large scale, where prior work is not so common. Certainly there are data sets, such as the "Parkinson's" dataset from UC Irvine's *Machine Learning Repository* [8], and the per-study datasets from many papers such as Ping et al. [9] and Gregory et al. [10]. Unfortunately, all such datasets that we could find suffered either/both from being 1) behind a paywall; or 2) too small for machine learning/classification research. Also, waveform access to audio samples such as we propose is not provided in any of the above papers or datasets.

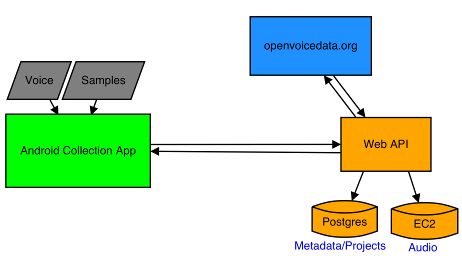
On the technical side, there is of course much prior work into mobile/web data collection, structured storage, and searchable availability. Our work depends on the research and open-source implementation of countless contributors in many computer science subfields: databases (both relational and NoSQL), web APIs, mobile application frameworks, web frameworks, and audio signal processing. We believe that by applying these technologies to an underserved area of interest in medical research and public health, and providing the results a free and publicly available forum, we are creating a novel project worthy of the previous technical/research work on which it depends. Furthermore, we hope that the OVDP can in turn serve as a platform for further research and implementation efforts in the classification, diagnosis, and treatment of dysphonias.

III. Project Narrative and Methodology

This semester, we have learned as much or more about process and the importance of domain knowledge as about the technology and implementation involved in our eventual deliverables. We began the semester with an interest in voice therapy on mobile devices for children on the autism spectrum, which in hindsight is an immensely complicated constellation of fields: mobile interface design/human factors for widely a varying target audience with special needs; voice disorders and therapy; child development; and abnormal psychology. While we learned much about these fields and enjoyed meeting active practitioners within them, our ultimate decision was that the year-long lab did not leave us enough time both to understand our problem domain and to implement a solution to address our lofty therapeutic goals. A particular difficulty was the data model for therapy sessions and the interaction it encapsulated: what was essential in a voice therapy session with an autistic youth versus optional and likely to be customized by individual practitioners? We found ourselves paralyzed at the stage of requirements analysis.

Eventually, we came to the conclusion that instead of arrogating (and performing badly) the roles of vocal specialist, child psychologist, and clinical therapist, we would concentrate on an aspect of the field that we discovered in our discussions with domain experts: active research in machine learning and classification of vocal dysphonias. We have already summarized our motivation to tend towards the infrastructural side of this problem in previous sections–once we isolated this domain, analysis and implementation proceeded much more quickly.

The architecture of our solution is as follows:



Each color roughly corresponds to a system component separable both by implementation/interface and by target audience. The Android app is intended for voice sample submissions from dysphonic patients (and control subjects); the API is for application programmers/researchers; and openvoicedata.org is for the general public, and researchers and patients interested in getting involved. This decoupled architecture allowed us to develop the various components in parallel; for instance, we worked with a stubbed-out API while developing the Android app until backend persistence was implemented.

The implementation technologies/frameworks are as follows:

- **API/Data Store** Go, MongoDB, Postgres, Heroku, Amazon EC2
- **Android App** Java, Android, Roboguice (dependency injection)
- **Website** HTML, PHP, Twitter Bootstrap, Heroku

We had previous experience in all of these technologies except for MongoDB and Android, so implementation proceeded without any major hiccups.

IV. Conclusion

This project has provided us with valuable experience in all phases of software engineering, from our protracted, multi-stage requirements phase through final integration of independently-developed system components. We hope that OVDP will prove to be of as much use to voice dysphonia researchers and patients as a community research tool as it has been to us as an exercise in larger-scale software development.

Acknowledgment

The authors would like to thank the following people for their assistance during our senior lab project:

- our instructors for EECS 581/582, Professor Arvin Agah and Patrick Clark
- Dr. James Khalili, Founder/Clinical Director, Lifeskills Management Center, Olathe, KS
- Dr. Jeffrey P. Searl, Associate Professor, Department of Hearing and Speech, KU Med
- Dr. Jonathan Brumberg, Assistant Professor, Department of Speech-Language-Hearing, KU
- Dr. Jane Wegner, Clinical Professor/Director, Department of Speech-Language-Hearing, KU
- Stephanie Meehan, MA, Department of Speech-Language-Hearing, KU

References

[1] L. Baghai-Ravary and S. Beet, *Automatic Speech Signal Analysis for Clinical Diagnosis and Assessment of Speech Disorders*, ser. SpringerBriefs in Electrical and Computer Engineering. Springer, 2013. [Online]. Available: http://books.google.com/books?id=yskKgs4RsKkC

[2] A. Tsanas, M. Little, P. McSharry, J. Spielman, and L. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 5, pp. 1264–1271, 2012.

[3] M. A. Little, D. A. Costello, and M. L. Harries, "Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures," *Journal of Voice*, vol. 25, no. 1, pp. 21–31, 2011.

[4] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity," *Journal of The Royal Society Interface*, vol. 8, no. 59, pp. 842–855, 2011.

[5] ——, "Enhanced classical dysphonia measures and sparse regression for telemonitoring of parkinson's disease progression," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 594–597.

[6] C. Sakar and O. Kursun, "Telediagnosis of parkinson's disease using measurements of dysphonia," *Journal of Medical Systems*, vol. 34, no. 4, pp. 591–599, 2010. [Online]. Available: http://dx.doi.org/10.1007/s10916-009-9272-y

[7] D. Voigt, M. Döllinger, T. Braunschweig, A. Yang, U. Eysholdt, and J. Lohscheller, "Classification of functional voice disorders based on phonovibrograms," *Artif. Intell. Med.*, vol. 49, no. 1, pp. 51–59, May 2010. [Online]. Available: http://dx.doi.org/10.1016/j.artmed.2010.01.001

[8] (2013, May) Uci machine learning repository. Web. University of California, Irvine. [Online]. Available: http://archive.ics.uci.edu/ml/

[9] P. Yu, R. Garrel, R. Nicollas, M. Ouaknine, and A. Giovanni, "Objective voice analysis in dysphonic patients: new data including nonlinear measurements," *Folia Phoniatrica et Logopaedica*, vol. 59, no. 1, pp. 20–30, 2006.

[10] N. D. Gregory, S. Chandran, D. Lurie, and R. T. Sataloff, "Voice disorders in the elderly," *Journal of Voice*, vol. 26, no. 2, pp. 254–258, 2012.

INFORMATION & TELECOMMUNICATION TECHNOLOGY CENTER
The University of Kansas

DEPARTMENT OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

THE UNIVERSITY OF KANSAS

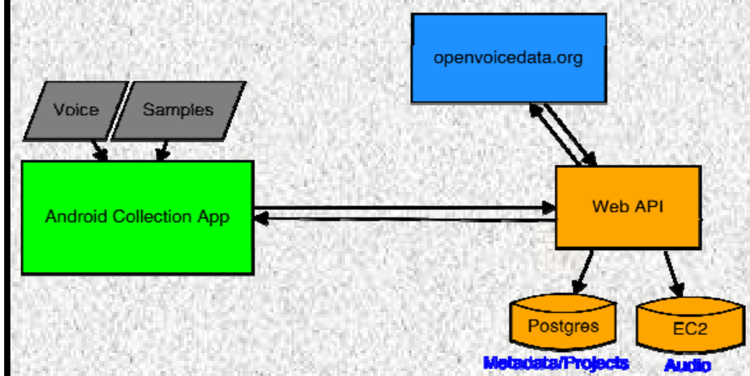
EECS 581/582: Computer Science Design

Lizzy Alonzi, Corbin Charpentier, Travis Cline, Patrick Flor

Problem Space / Motivation

- Medical data necessary for computational voice dysphonia research is not accessible enough
  - Modern approaches to research, classification, and diagnosis of vocal dysphonias are centered around machine learning and data mining; thus they require large, consistently structured data sets.
  - However, many current databases are proprietary, isolated, and/or not publically-accessible.
- Our goal: to provide infrastructure for the collection, hosting, and dissemination of open, publically-accessible dysphonia data sets, the **Open Voice Data Project**.

Architecture



Features

- A MongoDB (audio blobs) and Postgres (projects/data/metadata) backed REST API, capable of supporting research data requests and future applications
- An Android app capable of collecting voice samples according to structured dysphonia-specific project specs
- A website providing more information about the OVDP and displaying project statistics from the API

Future Work

- An iOS sample collection app to broaden the network of potential contributors
- Inlined DSP to calculate various characteristic features of waveforms
- Interactive data exploration/visualization facilities via the OVDP website
- Anonymized longitudinal data gathering, so participants might track their progress
- Data sets for other dysphonias besides Parkinson’s