Web-based Computerized Assessment of Speech Pathology Disorders

Steven Claggett, Vangelis Metsis :: Texas State University – San Marcos

{svc8, vmetsis}@txstate.edu

Research Supported by NSF Grant Number #1358939

Abstract

Early detection of speech-related disorders is of major importance in order for intervention and treatment to be effective. This work presents our effort to create a web-based computerized speech pathology assessment tool that users can easily access on-line in order to get an initial assessment of possible speech-related problems and seek further help.

Our system uses technologies such as HTML5, JavaScript, AJAX and Play! Web-Application Framework to provide an interactive front-end that can capture audio and video from the user's browser and also uses back-end technologies such as Java and MatLab to implement algorithms that analyze the captured data and perform the assessment.

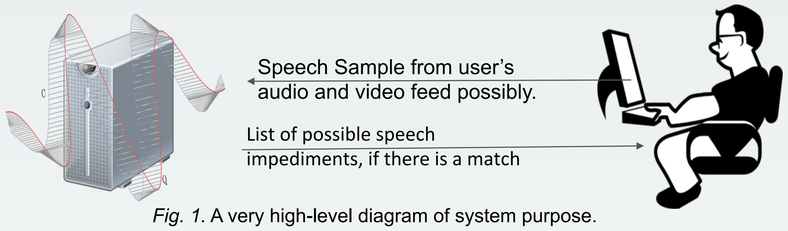


Fig. 1. A very high-level diagram of system purpose.

Introduction

According to the National Institute of Deafness and other Communication Disorders (NIDCD), approximately 7.5 million people in the United States have a speech or voice disorder [1].

While many disorders - such as dysarthria - are quite obvious when one listens to an affected speaker, many disorders - such as cluttering - are more mild in severity and may go undiagnosed, even into adulthood.

There is a lack of user-friendly applications which would aid to help identify possible speech defects that a user may have by using a speech sample provided by the user. We are attempting to create such an application which would be accessible via the internet.

Such an application would not provide a definitive diagnosis for a disorder. However, the application would bring the consideration to the user that he or she may have a speech disorder and should seek an official diagnosis. This application could also relieve pathologists from certain repetitive screening tasks.

Attempts at classifying speech into a small set of disorders have been relatively successful [2][4] using certain machine-learning and "in-house" algorithms. In order to measure dissimilarity between normal and abnormal speech patterns, our initial approach uses Dynamic Time Warping, both for its simplicity and its relative success at classifying certain types of time-dependent sequences [3].

Methods

In the early stages of the project, the main focus is getting a user interface that performs the functions of recording a user voice sample and uploading the file in order to be processed.

The front-end of the application involves the use of the user's microphone. The user is asked to repeat several recordings that are presented to him or her which are selected to exemplify certain traits of common speech disorders. The user records his or her attempt at replicating the sample recordings and the user's samples are then processed.

Using MatLab for the processing of the signals allows us to use algorithms which take advantage of the more expressive MatLab language and library. Fig. 3 shows an overview of our system.

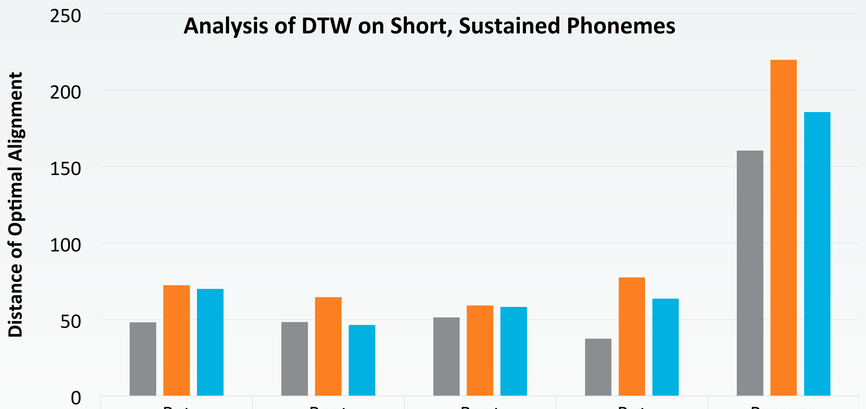
Application Features:
- Uses library to invoke instances of MatLab
- Invocations of MatLab run concurrently.
- Underlying Algorithm is "plug-and-play"
- Front-end uses HTML 5 features for video.
- GetUserMedia to capture media.
- MediaStreamRecorder to record streams.
- AJAX and XMLHttpRequest
- No page reload or loading.

Play Framework Features:
- Java & Scala based programming
- RESTful and asynchronous transactions
- State is preserved by use of cookies.
- Each user is provided a UUID.

Results

Our application uses Dynamic Time Warping to measure dissimilarities between normal and abnormal utterances of certain phonemes. DTW is also well-known for its ability to find an optimal alignment between two time-dependent sequences. The experimental results (Fig. 2) confirm our initial assumption that matching a normal and an abnormal utterance of the same sound would yield a higher distance than compared to matching two normal utterances.

DTW may be useful for detecting abnormality in speech, however, since the distance values for the *hoarse* and *flutter* voice qualities are very similar, it may not be suitable for detecting the specific type of abnormality. More advanced classification methods can be used for that purpose.



	Bat	Beet	Boot	But	Passage
Normal	48.1	48.3	51.3	37.5	160.6
Hoarse	72.5	64.6	59.2	77.6	219.9
Flutter	70.1	46.4	58.3	63.6	185.7

Fig. 2. A graph showing the result of DTW. Each series indicates the word sustained (for the "passage" series the phrase is "When the sunlight strikes raindrops in the air") and bar color indicates the type of voice abnormality. The "Normal" type shows the DTW distance when comparing two samples of normal speech whereas "Hoarse" and "Flutter" show the distances of the corresponding speech abnormality from the normal speech sample.

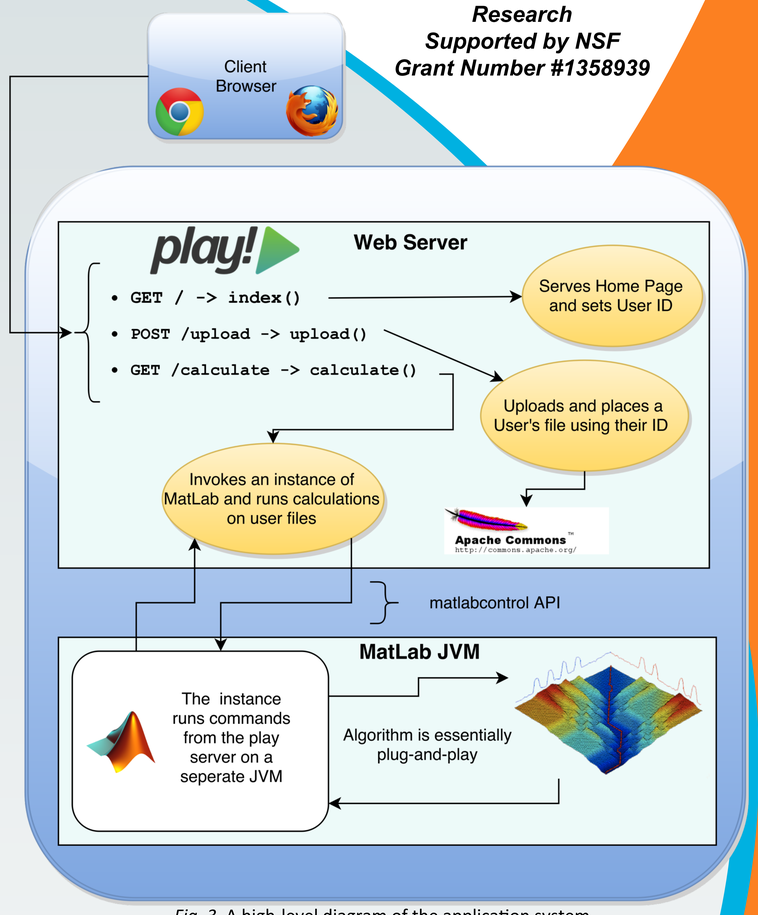


Fig. 3. A high-level diagram of the application system.

Future Work

Firstly, an analysis of suitability of algorithms must be completed. This will involve other algorithms that are suited for classification – such as Neural Networks and Hidden Markov Models.

Also, the system will have to be tested for stability. Currently, no more than a few active connections have been tested at a given time. If the system is to hold up to real-world use, it must be able to efficiently use its resources and coordinate many instances of JVMs.

Once we have a system that can perform the task to a sufficient degree, we will then begin subject testing and analysis. This will be a major factor in the efficacy of the system and will be the major result that will determine the usefulness of the system.

References

[1] Doty, R., Gates, G., Tomblin, B., Boyle, C., & Cruickshanks, K. (2010, June 7). Statistics on Voice, Speech, and Language. Retrieved July 28, 2015, from http://www.nidcd.nih.gov/health/statistics/pages/vsl.aspx

[2] Little, M., McSharry, P., Moroz, I., & Roberts, S. (2006, May). Nonlinear, biophysically-informed speech pathology detection. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* (Vol. 2, pp. II-II). IEEE.

[3] Metsis, V., Galatas, G., Papangelis, A., Kosmopoulos, D., & Makedon, F. (2011, May). Recognition of sleep patterns using a bed pressure mat. In*Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments* (p. 9). ACM.

[4] Martinez, C. E., & Rufiner, H. L. (2000). Acoustic analysis of speech for detection of laryngeal pathologies. In *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE* (Vol. 3, pp.2369-2372). IEEE

https://www.ummchealth.com/uploadedImages/UMHC/child_voice_speech.jpg