

LingoPros



Requirements Document

December 5, 2017

Josh Shaffer, Luis Montes, Erik Strauss, Matt Quintana

Sponsors: Dr. Okim Kang & Dr. David O. Johnson

Mentor: Ana Paula Chaves Steinmacher



Client Name: _____

Client Signature: _____

Overview: The requirements document will detail the different requirements that we have discussed with our client in order to implement the project to their satisfaction. It will also display the problem statement, our solution vision, and our plan for development.

Table of Contents

Introduction	2
Problem Statement	3
Solution Vision	4
Project Requirements	6
Functional Requirements (FR)	6
Performance Requirements (PR)	9
Environmental Requirements(ER)	10
Potential Risks	11
Project Plan	14
Conclusion	15

Introduction

Non-native English speakers often have a hard time articulating their speech, not just in what they say, but how they say it. Proper emphasis on words or just general speech tone is something that native English speakers do not have to think about in their everyday speech as it is something that just comes naturally. For example, in a conversation between a non-native English speaker and a native speaker, the non-native speaker may not be aware of how, by not matching the tone or cadence of the native speaker, the conversation can begin to feel awkward and unnatural. Recognizing these differences in tone, pitch, stress and other prosodic features between native and non-native English speakers, is one of the key focuses of Dr. Okim Kang's research at the Applied Linguistics Speech Lab (ALSL) at NAU. She has created a program that is able to analyze recordings of non-native English speakers and grade their English proficiency based on aspects like tone and emphasis compared to native speakers. However the current system is ran locally on her computer, instead she wants an easily accessible web portal for users to upload and analyze audio samples in one location.

Currently, students working with Dr. Kang at the ALSL have to do all of the speech analysis by hand, listening to audio samples and recording what features they hear or recognize. This process can take hours on end even for a small 10 second audio clip, so as such, an automated process of speech analysis that is easily available to the students would drastically improve their research capabilities by saving them time and energy. Dr. Kang has already developed her own experimental speech analysis program that can recognize prosodic features, however it is cumbersome to use as it is spread across two different machines and operating systems. Additionally, the model that the current program uses, a framework based on a distinct form of speech analysis posed by David Brazil called *discourse intonation*,[1] is not recognized as the standard in speech analysis. A set of conventions called Tones and Break Indices (ToBI) is considered to be the standard for annotating speech features, and so Dr. Kang wants to show that the Brazil program performs just as well or better than automatic analyzers based on ToBI.

We have been brought on to develop a speech analysis program based on the ToBI model and implement the web portal that will house her experimental program based on the Brazil

model. The first part of our project involves using a third-party speech analyzer called AuToBI to extract prosodic features from speech samples. We will then develop a machine learning program to score those extracted features, using the output of the AuToBI program as an input. The point of using machine learning is that it can better determine a more accurate score that is consistent with human scoring. Once the speech analyzer is built, we will then move on to developing the web application to house Dr. Kang's framework (David Brazil Model) so that her program can be accessed from anyone, anywhere, assuming they are an NAU student and Okim allows them to have access to the site. In this document, we will outline the requirements our client has laid out for us and display the problem and our solution.

Problem Statement

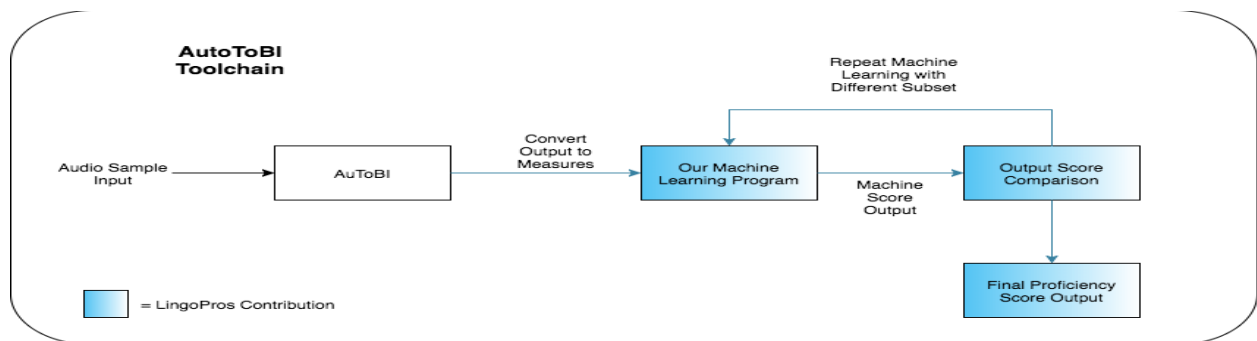
Of all the issues that we have discussed with Dr. Kang, we have narrowed the scope of the project to be something that we can accomplish in the given time frame. There are really three main problems with their current workflow: the David Brazil model is slow to use, it is not in a convenient place where students can run the program, and it is criticized for not following the industry standard ToBI model. These three problems have all contributed to making a negative impact on Dr. Kang's workflow and, as such, these are the three problems that we need to fix. After fixing these three issues, most of their current problems will be solved and they will be in a position to continue advancing their research in a timely manner.

Solution Vision

Our solution is essentially split into two parts. The first part, and the bulk of the work, is developing a speech analysis software that uses the ToBI framework. This will be used to compare with the David Brazil Model and determine whether it is actually better than the ToBI version. The second part is developing a website that will host the David Brazil Model for approved users to access and streamline the result gathering process for our users. We will first establish an application web server that can run the MATLAB-based David Brazil Model analysis program given to us, hosted on a server provided to us by the NAU ITS department. Second, we will create a website with a user interface to upload files to the the hosted program. This will solve the speed issue by having everything in one place as well as solving the accessibility issue by allowing anyone to access it online.

Specific details of our solution are listed below:

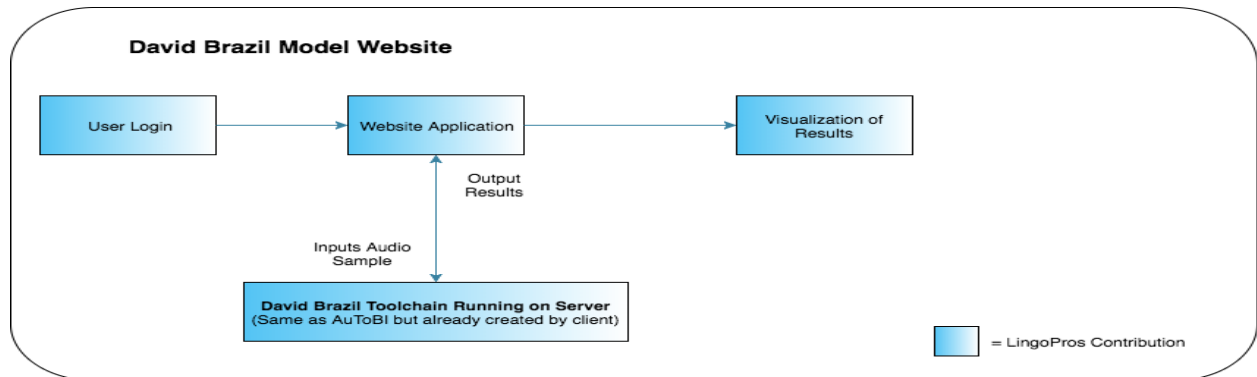
Part 1: AuToBI Prosodic Labeler



The above graph is a rough outline of our solution for the prosodic labeler that we will be designing. The user will upload an audio file into the AuToBI program. AuToBI will perform its analysis and return the measurements of different prosodic features of the audio file. Those measurements will then be passed into our machine learning program which will perform an analysis based on the measurements and prepare an output score. Before it finalizes the score, it will compare the score to that of a human's score of the same file. If the score is not close enough, the program will run the analysis again with a different set of measurements. It will

repeat until it can determine which measurements best contributed to an estimated score that is close to the human score. At that point, it will return the results of the analysis to the user.

Part 2: Host David Brazil Online



The above graph outlines how the David Brazil Model will operate within the online server. Users will first login to their account. Users will not have full access unless they have been approved by Dr. Kang. Once the user is logged in, they will be able to upload an audio file straight into their browser and the analysis will be performed by the Brazil program on the Tomcat server. Once the analysis is complete, the user will be redirected to a new page which will show information about the file and the results of the analysis. This information will be stored in each user's account so they can reference these results without having to run the analysis again.

Project Requirements

We have discussed the project heavily with Dr. Kang and we have determined the core requirements necessary for our project to be successful. The agreed upon requirements are all listed below in three categories. Functional requirements are items that users must be able to do/use. Performance requirements are requirements our program must meet in terms of performance (speed, accuracy, etc.). Lastly, environmental requirements are specific constraints imposed on us that we are unable to change.

Functional Requirements (FR)

Web portal

“As a linguistics researcher (or linguistics student), I want to be able to go to a website and upload sound files I have created (or been given) to a service from that site so I can get back prosodic results based on the David Brazil model without needing to install and run the full application on my local machine.” - a user’s story

FR1: Upload Audio Files on Website

On the analysis page of the website there will be an upload file button that brings up a file explorer window to select the .wav audio file to be uploaded for analysis. The user will then click on a run analysis button to pass the file from the web server to the Tomcat server. The sending of the sound file to the server is achieved by an AJAX post request of the input sound file to the server’s running Brazil program which will analyze the sound file for different linguistic features. After the program finishes its analysis the server automatically sends its own AJAX post request to the client’s browser to present resulting data in the form of HTML with tables. The results consisting of prosodic events and anomalies that the David Brazil Model determined are sent back to the user’s browser window to be displayed.

FR2: User Login

From the homepage of the website, there will be a login option for users to click on and enter their username and password. Once logged in, the website will display an area to see results from their previous analyses and link to go to a separate page on the site to run a new analysis.

FR3: Retrieve User Profile and Results

When a user logs in, a query will be sent to the SQL database to check if they are a valid user and if they have entered the correct password. On valid entry, a new query will be made to return user information for display such as username and records pertaining to previous analyses.

FR4: Create User Profile

For users without a profile, on the homepage of the website there will be a signup option where the new user will be taken to a page where they can enter their information into fields regarding name, email address, username and password. Once their information is entered into the fields, the user will click on a registration button which will send a query to the SQL database to create the new user record.

FR5: Reset password

From the login screen on the website, if the user has forgotten their password there will be an option for the user to reset their password. Once they click on the reset option, a message will be sent to the email address that they originally provided on registration with a link for them to reset their password.

FR6: User Login History

“As a linguistics researcher (or linguistics student), I want to see my previous file analyses from the LingoPro’s application after I log in . I should be able to just click somewhere to view my upload history so that I can reinvestigate my research without having to reupload sound files that were analyzed and some point prior and waste time

waiting for the hosted David Brazil model program to finish calculating previously seen results.” - a user’s story

Our solution for users to avoid reanalyzing files they may need another analytical inspection at is to keep their prosodic data in the same database that also holds their user accounts. This way, our website can query our MySQL database for a particular user’s past upload results with their account being the key to finding that data. The database will also need to be queried upon completion of analysis to store result data to the account that started the analysis through uploading.

On the home page for a logged in user, the user may press a button labeled “My History”. Upon pressing such button, a new HTML rendered view will appear that shows results of previous data analyses he or she has entered.

AuToBI based Prosodic Labeler

FR7: Audio File Input to ToBI Based Analyzer

For the prosodic labeler that our team will develop, the program will take in a speech sample and pass it to the AuToBI program for analysis. AuToBI will analyze the speech file for various linguistic features and then produce an .arff file, an output file, and command line results that can be used in the machine learning analysis. The .arff file will contain a list of attributes and several rows of records of data. Each record will have data pertaining to each of the attributes. These data values and corresponding attribute will be used in the machine learning component of the program.

FR8: Machine Learning Component

The results in the .arff file will be passed into a Java application that uses the Weka suite of machine learning software to run the process of feature selection on the results. Different subsets of attributes, or features, will be selected from the results, and those values will be used to estimate a proficiency score based on an evaluation function. Based on the difference between the estimated score and the actual human evaluated proficiency score, the weights on certain features will be lowered or raised based on how they contributed to the accuracy of the estimated

score compared to the human score. A new set of features will be selected and a new estimate will be created. The program will loop through this feature selection, weight adjustment, and estimation until the estimated score is within a small margin of error from the human score.

Performance Requirements (PR)

PR1: Scoring Accuracy

Involves FR5 and FR7

In order to come to produce an effective speech analyzer that is comparable to the one developed by Dr. Kang, our prosodic labeler must have a similar computer to human scoring accuracy as Dr. Kang's program. That is to say, the Brazil program currently has a 72% accuracy to human scorings when performing its analysis on audio samples, and so we would want to have our prosodic labeler reach at least 72% to 75% accuracy with human scorings in order to be adequately comparable. An accuracy rating can be obtained by running an analysis on large set of samples and then checking the difference in proficiency scores of each one to get an overall accuracy rating.

PR2: Analysis Runtime

Involves: FR1

For the web portal, the analysis has to run faster than both the manual analysis, which can take several hours to complete, and the Brazil analysis run over two machines, which can also take several hours to complete. By reducing the need to manually transfer results from one machine to another, we hope to reduce the full analysis runtime by at least 50%.

PR3: Usability

Involves FR1 and FR3

Because most of our users will be from the field of linguistics, they may not have as much technical experience, so we want to make their user experience and their workflow on the website, as smooth as possible. This means that main functionality such as upload and analysis should be directly available after initial login, at most one click away.

PR4: Code Security

Because the Brazil program and all of the code for it is copyrighted under our clients, we will have to protect the code from outside parties trying to copy the code for their own use. This means that none of the code or the functionality should be visible from the client side, and any changes to the Brazil program's codebase should not be made publicly available.

PR5: User Security

Involves FR2, FR4, and FR5

Although we will not be handling extremely sensitive information, we will still want to have a basic security system to protect our users information. This would involve sanitizing inputs in order to prevent SQL injections into our database that could compromise our users' info.

PR6: Readable Code

Because both our prosodic labeler and web portal will most likely be improved on by future capstone teams and looked over by non-technical individuals, the code needs to be easily readable and understandable for anyone to look at. The purpose of each piece of code needs to be clear just from reading comments and documentation, or through appropriate variable and function names that describe their purpose. Our aim would be to have all of our functions commented and to create a readme file for both the web portal and our prosodic labeler, explaining how each works.

Environmental Requirements(ER)

ER1: AuToBI

Involves FR7 and FR8

For the machine learning prosodic labeler portion of the project, we are tasked with developing it using the standard ToBI model of speech analysis. Rather than completely developing our own speech analyzer, we have to use an open source tool for automatic speech

analysis known as AuToBI, which is based on the ToBI model. Our program will have to be able to accept the results produced by AuToBI, and then also interpret them in a way so that the prosodic labeler can accurately compare the scores to human scores and display those results.

ER2: Previous MATLAB Implementation

Involves FR1-FR6

Dr. Kang and Dr. Johnson have asked us to host their previous Brazil program on the web so that their students at the ALSL can run the analysis from anywhere. A full analysis depends on two halves of the program, the first half in MATLAB and the second in Java, which means that we will have to work with the MATLAB implementations and convert them to Java classes so that they can run on our Tomcat server.

Potential Risks

This project has a series of risks associated with that involve three different parties. Ourselves, our client Dr. Kang, and the future users of our final product. The risks are rated on likelihood and severity. Likelihood is described from one meaning very unlikely to five meaning very likely. Severity is rated in similar magnitude of one to five, but with one meaning a benign threat to five meaning this group would have to be disbanded due to danger to some party or infeasibility.

- **To the Developers:**

Involves PR6

Legal Issues: The algorithm implementation of the David Brazil based model from Dr. Kang and Dr. Johnson is copyrighted. While our version control system is Git and we have a repository for the parts of the Capstone we write, we need to ensure our client's code is hidden so as to avoid the chance for anyone else to put her algorithm on a different application server and do the same processes we are trying to accomplish in part two of the project without proper credit and/pr payment to her and Dr. Johnson. We need

to hide the source code parts of the David Brazil model so as to make stealing of material impossible from the Internet alone and thwart heavy financial losses Dr. Johnson and Dr. Kang could take in that event. While the severity of this risks comes in at a stunning four, the likelihood rests at a one due to the difficult nature of putting the MATLAB code as a functioning program on an application server in the first place.

- **To the Client**

- Involves FR1-FR6**

- NAU Storage:** Provided we get our website application up and running, complete with a user system and interactivity with our application server, our database will be collecting a potentially large amount of data on sound file scoring and user account information.

- Since our database is hosted from NAU, the NAU ITS team may choose to not collect her and her user's data on the system when the database is full since space is limited. Dr. Kang and ITS will have to decide on how much free storage may be allotted to this project and what to do in the event of an overflow. We could make the system use as much free space as possible and indicate to her when it nears its max capacity. The likelihood of NAU ITS deleting data is a very low one rating since a small scholastic circle will be using our product and hence not filling to much of the NAU provided storage space and also . The risk, if it did actualize, is of a mild two rating because the user accounts are simple to to be re-registered and re-uploading files to get the same analyzed results again is also intuitive... although admittedly this potential problem of data loss is a pesky time consumer for the users.

- **To the Users**

- Involves FR1-FR6**

- Tomcat Crashing:** There is the possibility of our application server crashing. Tomcat servers usually crash due to memory usage problems and with a large amount of users to our system, this a server crash is a possibility. In the event of a server crash, the solution is to restart the server with a simple one line command from the terminal. That said,

users could not use our web application until the server is back online and that may make finishing their research or other work less likely to complete on time. The likelihood of this happening is a mildly low two rating because the program is as a whole not expected to be memory intensive to our server, but perhaps with more users the likelihood of crashes rises. The severity of a server crash to our client is a one because the server can be easily reset from the command line, but a severity rating of three is given to non-administrator users who would have to wait patiently until our client restarts the server with the one liner reset command to the terminal that the server was started from.

Database Leaks:

Involves PR5

In the event of our database being made public by a third party, we plan to limit the privacy breach of our users by B-Crypt hashing their passwords during the creation of their profiles. A database leak of our system would reveal insensitive data due to no stored passwords (only one way hashes) being present and the nature of scores from sound samples being inert in reputation or financial impact. The likelihood of a database leak is a low one rating because we will sanitize any and all of our website's text field inputs to avoid cross-scripting attacks. Also our database is not an attractive target for mischievous hackers because the data stored therein is not financial related, and even if someone unauthorized to see our database actually saw our database, they'd see no passwords stored that they could've used for more mischief to our users. That leads to the next point that this low likelihood event happening has only a severity rating of one because only insensitive data would exist in our database.

Project Plan

AuToBI

For the machine learning portion of the project we will be creating demos for the different pieces and making sure each work desirably before combining them into one cohesive unit. The different pieces are converting the AuToBI output ARFF file, the machine learning functions, and decision trees which will all be demoed separately. Once we combine them we will create a functional prototype so that we can show it to our client and make any changes they want.

David Brazil Website

The website will also be split up into different portions with regards to demos to make sure each piece works before putting them all together. These pieces will be the server hosting MATLAB programs, the website/database, and the server communicating with the website. Once they all work we will begin creating a prototype of the David Brazil program so that it runs on the server and sends the output to the website for users to see. Below is our tentative timeline for production:

Task Name	2018												
	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	
Semester 1	Semester 1												
1 Team Standards and Team Inventory													
2 Technological Feasibility													
3 Requirements Document													
4 AuToBI output conversion demo													
5 Machine learning program demo													
6 Tomcat server hosting MATLAB demo													
7 Website, server, and database demo													
Semester 2	Semester 2												
8 AuToBI output input to machine learning program													
9 Tomcat Server hosting David Brazil program													
10 Machine learning program prototype													
11 Website with David Brazil program prototype													
12 Extensive Testing and Updating													
13 Finished Product													

Conclusion

Automatic speech recognition is a topic that has been around for several decades, however it is only recently that the field has advanced to the point where researchers are able to study the deeper subtleties of speech due to advancements in computational capability and machine learning. Linguistics researchers are now taking advantage of these new, better performing tools to carry out their analysis so that they can gain a better understanding of how humans communicate. Dr. Okim Kang is one such researcher who is taking advantage of these new technologies and using them to develop a new framework for automatic speech analysis.

Dr. Kang has asked our team to help develop a speech analyzer based on the standard speech analysis framework ToBI so that she may compare the results against her Brazil framework. In addition we have been asked to build a web application that can provide our client and other users a website to analyze speech sample using this Brazil framework that our clients created. The current system they use is cumbersome and time wasting because they have the framework spread across different computers with specific operating system installations and an arduous build process per each installation. Our solution will help Dr. Kang and her students at the ALSL conduct better research so that the field of speech analysis as a whole can gain a better understanding of how humans communicate.

For our machine learning solution we plan on making sure that BigML correctly works with the ARFF output file from AuToBI by using Weka to convert it to the proper format with the suprasegmental features. We believe we will be able to meet or surpass our client's expectations throughout the development of this project, and hopefully provide them with a useful speech analysis tool for years to come.