Athena: Final Project Proposal

Laura Bochenek, Fangzhou Liu, Mounica Kota, Jing Sun, Florencia Spinella

1.The problem

English is the main language in the world. There are 1.2 million international students in the US, all of whom, on average, got the lowest scores in their Speaking section of the Test Of English as a Foreign Language (TOEFL)[1]. Not being understood makes it hard for students to reach the audience while presenting, discourage them from class participation and acts as an obstacle for social integration. Athena will help them overcome this obstacles by providing a quick way to let them say what they want to say in the right way, that is, with the correct English pronunciation.

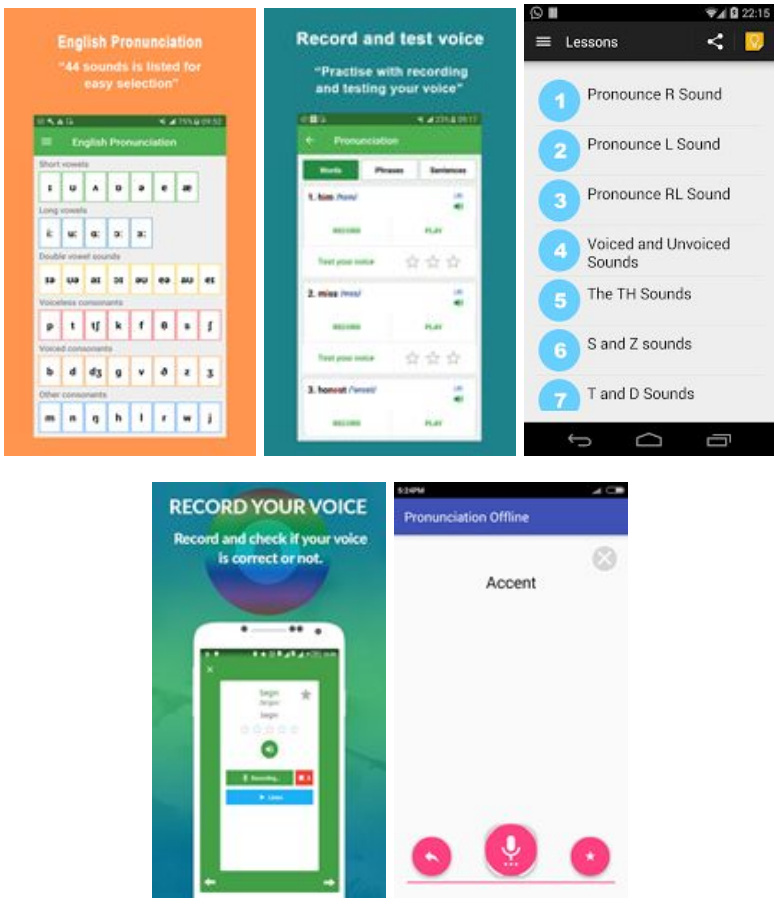
Not only are we motivated by 1) some of the group members own experience to struggle with English pronunciation while being in the US, even though we have studied English for +7 years; 2) Students we know from our own classes who have a hard time when they have to talk or present (such as our first HCI class where all of the students had to introduce themselves); 3) Having witnessed presentations where we could not understand what the person presenting was saying and therefore, we lost interest: 4) But our belief is that this application can have a greater purpose. Research shows us there is a potential use to help people looking for a job, indicating that *how to speak English is more important than having papers* for immigrants in the US and that *people whose words cannot be understood look less smart* for Americans. Additionally, research shows that *immigrants get substandard health care* because they do not speak English well and only a few hospitals have translators[2].

Therefore, while our currently target user is US international students we envision our technology could be, without having to add many more features, targeted to help immigrant users to get a job and improve health service they receive.

[1] https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf
[2] http://www.cbsnews.com/news/language-barriers-cause-problems/

2. The solution

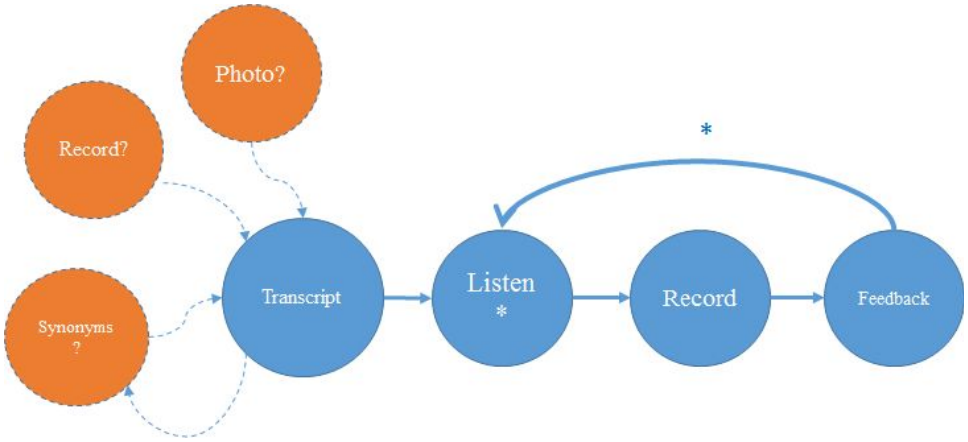
Today, there is plenty of mobile applications and software intended to train the users in English pronunciation. However, their approach is different. Instead of focusing on user intended content pronunciation, these applications are structured in an English course manner, where the user goes through different levels of knowledge and evolves through time. They offer an academic approach where each sound is to be practised based on fixed words that are provided. There are tools that outperform others, and those are normally paid. While the pronunciation feedback is present for free text that the user input, that is normally one of the complementary features of the app and users might not even know how to get there, or even will not even download the app because of the description and the difference in the app approach and the approach they are looking for, which we believe is a fast, straightforward feedback on their speech pronunciation.



Samples of current applications available

We will provide that fast, straightforward feedback on users speech pronunciation. We envision an application that will obtain the text the user wants to pronounce, will provide the correct pronunciation, user will record and a feedback will be provided. We will get details on feedback as we perform interviews, but we imagine that the text will be presented to the user with colors indicating whether the word was pronounced correctly: from green if pronunciation is accurate to red if the word was mispronounced. For those mispronounced, the user will be able to click on them to listen the right pronunciation and practice to record again and get feedback (check our video!); so on till the user is satisfied or the word is pronounced correctly. Additionally, a percentage score will be displayed. For more details on this score, please check out the implementation section.

How the application will obtain the text is another objective of the needfinding we will perform. We believe typing and recording might be the most useful methods, but we want to get that information from users. We also believe other features such as synonym finding based on input speech might be another nice feature to have. It would help people who lack vocabulary.



Athena workflow

3. Needfinding

We will be using two needfinding techniques in order to better understand our customers and target market. Specifically, they are **online surveys, and personal interviews**.

We chose online surveys so that we could get answers to general questions related to our problem and product, for a mass audience. It would allow us to better gauge who exactly our target customer is through a large pool of data.

We will be posting the survey through multiple online mediums. Four of the five group members are international students, and this gives us a large network of possible customers. We will be posting on each of our individual facebook accounts to attract users of multiple nationalities. Along with this, we will be posting the survey on whatsapp groups in which we feel potential customers could exist. A few that come to mind right now are friends groups from our home countries, along with professional groups where English was not the main medium of communication.

We chose to go with personal interviews for a few reasons. One is because we can ask more open ended questions in order to draw pertinent information, and can always ask a follow up question which could not be done in an online survey. Another advantage of the personal interview is that we can be more selective of the participants. This way, we can choose the participants which we feel would benefit most from our product, and through an iterative process, become more accurate in the said selection of participants.
We will select candidates who we feel are our target market; specifically international students with low exposure to the english language prior to coming to the United States.
We will conduct the interview using the following protocol:
Introduction, Kickoff process, Build rapport, Main experiment, Reflection, Wrap up

<u>Online Survey Questions:</u>

1. What is your background?
   (Graduate student, undergraduate student, PhD Student, Others)
2. What country are you from? Is English your mother language?
3. What do you think of your pronunciation of words?
   (Perfect, Pretty good, Well, Pretty bad, Very bad)
4. Do you have a hard time understanding others when they speak English?
   (Yes, No, Maybe)
5. Do you use electronic dictionaries to check words?
   (Yes, No, Maybe)
6. How often do you have to give presentations for class?
   (Very often, Ofte, Sometimes, Pretty rare, Hardly ever)

7. What kind of feedback do you usually get in class from your classmates on your presentations?
(Very good, Pretty well, Good, Pretty bad, Awful)

8. What do you think of those existing tools in the market right now which can help people with their pronunciation skills in English?
(Very helpful, Pretty helpful, Good, Pretty bad, Not at all helpful)

9. Why?

Interview Questions:

1. Do you have trouble pronouncing words in the English language?
2. What are some situations in which you faced difficulties with pronunciation?
3. Did you do anything in the future to overcome such difficulties?
4. Do you feel people have a hard time understanding what you have to say?
5. Do you have a hard time understanding what others have to say in English?
6. Do you believe there are tools in the market right now which can help people with their pronunciation skills in English?
7. What method do you usually use while preparing for a presentation to help your pronunciation?
8. What dictionary do you usually refer to to check your pronunciation? Why is that?
9. Do you think speaking words right helps you with communicating with people?
10. Where are you from? Do you think people from your country has some problem with pronunciation?

4. Prototyping

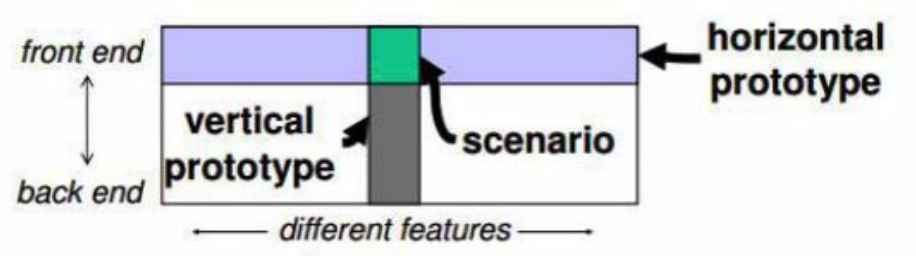
For prototyping, we will be implementing three prototyping techniques. Specifically, they are **storyboarding, paper prototyping, and wireframes**.

We will be using storyboarding in order to show what type of situations would merit using our app/website. The results of the needfinding will allow to show the personas and situations under which these personas would use the app.

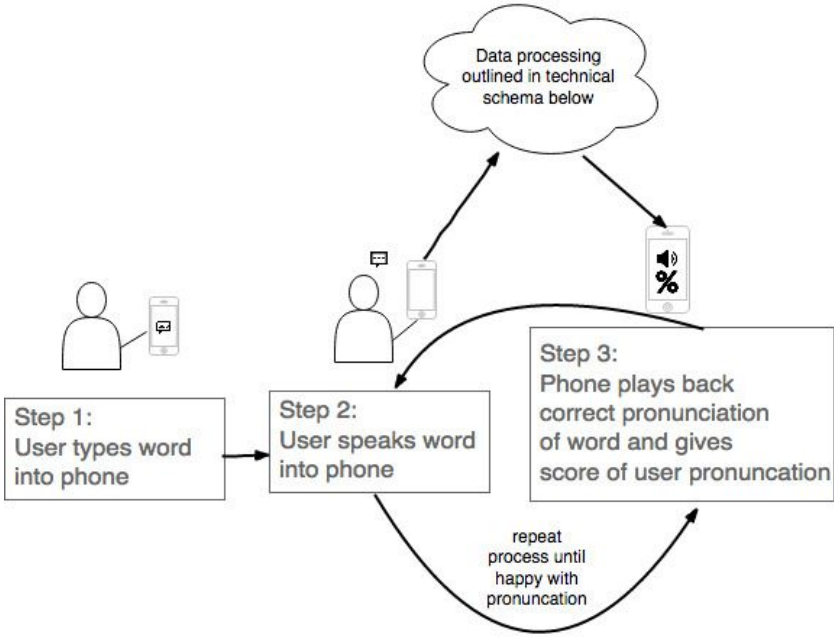
We will go with paper prototyping in order to get a better understanding of all the functionalities the app would incorporate without the huge loss in time that would exist in an actual prototype.

Finally, we will create a wireframe to better visualize the look and feel of the app. In order to test the validity of the paper prototype and the wireframe, we will use the Wizard of Oz method, where the “wizard” would be an impartial candidate.

We would run each prototype through several iterations, where there would be an increment in breadth and fidelity with each iteration.



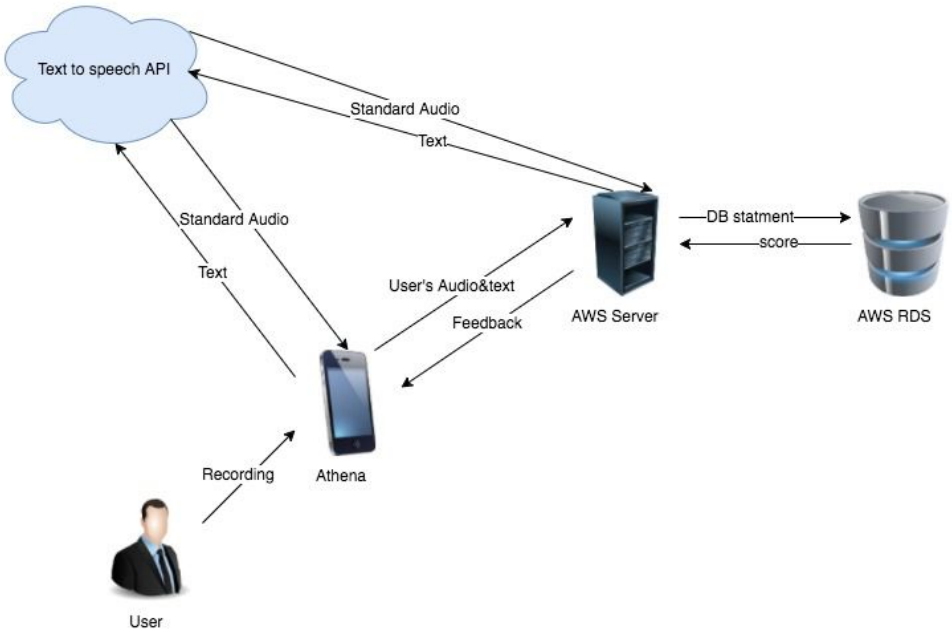
5. Implementation



Athena is going to have features that will allow users to improve their English pronunciation. With Athena, the user will have an user account that will keep track of their pronunciation progress. The user will be able to record themselves speaking a word in English, and then type in that word to the app. The app will then play back the proper pronunciation of the word and provide a score of how close their pronunciation was to the proper pronunciations.

Athena is going to be built as a mobile app. We believe that Athena will be most useful to users as a mobile app that they can carry around and use on-the-go as needed. While we believe that a mobile app is the best choice for our users, this choice comes with the certain drawbacks. Because the processing power on mobile applications is limited at best, our implementation will make use of a server to handle the pronunciation comparison algorithm. This means there will be a lot of data transferred back and forth between the server and the mobile app, and bad network connection could affect the speed and usability of the app.

Athena is going to be written as an iOS app using Swift and Xcode, largely because some team members have experience in the iOS coding environment. This frontend will handle the user interaction of both creating and managing their accounts, and recording and inputting the text of their words, and the displaying of the proper pronunciation and the score of the user pronunciation. The backend is going to be a server hosted by Amazon Web Services (AWS). Amazon's API Gateway will be used to communicate information between the iOS app and the server, AWS Lambda will handle the API requests from the iOS app and to the text-to-speech API, and run the pronunciation comparison algorithm, as well as store user account information in the AWS Relational Database Service (RDS). AWS was chosen because it offers all of our required services on a free tier level. AWS Lambda will be run using the Python SDK on the backend, as Python has many open source libraries such as LibROSA and Dejavu for audio processing and analysis. Additionally, a third-party text-to-speech API, Voice RSS, will be used to get the correct pronunciation of the user's target words.



Athena faces many technical challenges. The largest challenge is creating a meaningful comparison between the correct pronunciation of the words and the user's pronunciation of the words. To solve this problem, compare the frequencies of the user pronunciation and the proper pronunciation. First, we will do a Fourier Transform using the FFT algorithm, and then we will compare the frequency distribution. This will allow us to compare the two audio files without the volume and the frequencies themselves affecting the comparison. We additionally plan to use the Mel-Frequency Cepstrum Coefficient Algorithm (MFCC). Mel-Frequency Cepstrum Coefficient Algorithm is a popular algorithm designed for voice recognition. In general, it use a nonlinear scale(Mel Scale), which can better simulate our hearing system, to represent the voice signal. To get the Mel-Frequency, following steps are required:

- Split the audio signal into a series of frame.
- Pre-emphasis the signal and then filter out all noises via a HPF(High Pass Filter).
- Transfer the signal from time-domain into frequency-domain using Fourier Transform.
- Transfer the scale from frequency to Mel-scale using Mel Filter.
- Take the logarithm of all filterbank energies.[3]

[3] Mel Frequency Cepstral Coefficient(MFCC) tuturial, 2013, http://www.practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/

- Apply IDFT(Inverse Discrete Fourier Transform) towards each frame, then we got the cepstrum of this signal.
- The amplitudes we got is the MFCC and the first 13 coefficients are necessary for voice matching.

Given these two audio comparison algorithms, we plan to deploy two scoring schema respectively.

- For frequency distribution algorithm, we plan to use KS Test (Kolmogorov-Smirnov Test). In KS Test, the maximum distance between two samples will be returned as the difference between these two samples. We can use this difference as our final score. The shorter this distance is, the better score you will get.[4]
- For MFCC algorithm. We can simply compare whether the MFCC coefficient array in user's audio is the same that in standard audio. But, we need to do further experiment to figure out the weights of these 13 coefficient. The final score can be calculated by $score = \sum_{i=1}^{13} \Delta x_i \cdot \omega_i$.

Another challenge is getting a library of the correct pronunciations of words. To accomplish this, we decided to use to text-to-speech API that takes in text, and then sends back an audio file of a computerized pronunciation of the text. Specifically, we plan to the use previously mentioned Voice RSS API, which allows 350 free requests a month. A potential issue with this solution is if the computer pronunciation does not sound accurate enough to real life, and the comparisons with the user voice is inaccurate. If we find the comparisons to be inaccurate, we plan to adjust the algorithm to account for these differences. We looked into using a pronunciation database, but there are not any that are free, and the cheapest one does not distinguish between American and British English, so we decided that we will need to stick with the Voice RSS API.

Other small challenges include recording the user's audio in the iOS app and sending that audio to AWS for processing. To record user audio, we plan to use the AVAudioRecorder iOS API. To send the audio file to the server, we will use the

[4] Kolmogorov-Smirnov Goodness-of-Fit Test, Engineering Statistics Handbook 1.3.5.16, http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

NSMutableData API, which wraps files to allow them to be sent via byte buffer to the server.

6. User Study / Evaluation

Our hypothesis while conducting evaluation studies is that the users who use the Athena app will benefit the most in their pronunciation skills, compared to users who do not use any tools as well as users who using competitive apps in the market.

The experiment will be conducted as follows: Each participant would be given a paragraph to read. The participant would first say the paragraph without prior practice, then take 10-15 minutes of time in order to learn the paragraph better using various mediums outlined further. After the time is up, the participant will again read the paragraph. An impartial judge would determine the improvement in pronunciation of the participant. The **metrics** we will be using in order to determine the result of the experiment are understandability of participant, ease of use of app, and user ratings.

Experiment Guidelines:

There will be 15 participants who are divided into 3 different groups.We will choose individuals who range from average speaking to poor skills in English. These individuals will be from various ethnic and cultural backgrounds in equal proportions in all the 3 groups so that error is minimized.

One of the three groups is the control group where the participants do not use any tools or techniques to help improve their pronunciation skills. The second group will use Google in order to better their pronunciation. The third group will be using Athena in order to help improve their pronunciation. After the experiment is run, we will test to see which group saw the most improvement in their performance. This is a between subjects study. The factors which will remain constant are time spent on one project, the presentation itself, and day the experiment is conducted for participants.

Fallback Plan

1. If the program does not work out fine, we can do some evaluation survey to all those participants talking about their feeling after perform the three programs separately.
2. If the results for all participants set equal weight to our App and to the comparative App, we will just enlarge the population to do the survey again.

One feature which we propose is that a user's presentation should be able to be connected to the app so that pronunciation of words in the presentation can be taught to the user. Though it would be a good addition to the app, we feel that it is difficult to implement and hence can be justified as a risky feature. If, at the end, this feature does not work as intended, we are comfortable with removing this functionality. The reason behind doing so, is because this is not the main functionality of the app, it is simply an add on.

7. Timeline & deliverable

https://tinyurl.com/hu4jvph

8. Concept video

Enjoy at https://tinyurl.com/jcfe2yo

About Us

Team Member	Degree	Field of Study	Project Role	Past Experience	Skill Set
Florencia	Masters	TEAM	Project manager, app design, video & poster	Bachelor's in Systems Engineering; IT Project Manager for General Motors and Anheuser-Busch Inbev developing IT solutions for business users	Project Management, Photoshop, Illustrator, Video Editing, Python, C, SQL
Laura	Fifth Year Take 5 undergrad	CS, Psychology	App development	Prior full stack software engineering internship at Etsy	Javascript, Java, SQL, PHP, git, HTML
Fangzhou Liu	Masters	CS	Pronunciation comparison algorithm / app development	Bachelor's in Telecommunications Engineering; prior iOS software engineering intern at a startup in Beijing	Java, Objective-C, Ruby, SQL, Web and Mobile development
Sunny	Masters	Computer Science	Web development	Bachelor's in Information and Computing Science; worked as a web developer for one year	HTML, Javascript, jQuery, Objective-C, Java, SQL
Mounica	Masters	TEAM	App design, video & poster, needfinding and evaluation techniques, prototype creation	Bachelor's in Electronics and Instrumentation Engineering; worked as a systems engineer	SQL, Python, HTML, Excel, Wireframes, Poster creation