Speech to Text for Monitoring Nutritional Health

Josh Kolasch, Niloofar Hezarjaribi and Hassan Ghasemzadeh
School of Electrical Engineering and Computer Science, Washington State University

Abstract

Diet and physical activity are important for prevention and self-management of many chronic diseases. Self-monitoring is one of the earliest skills in nutrition management; yet, this technique is cumbersome, and adherence is low. Therefore, many researchers focused on automating this task using sensors or applications embedded in smart-phone devices. Mobile sensors such as an accelerometer have been used to detect eating moments and manage physical activity. This technology has low user adaptation and compliance. Mobile application requires users to enter text or take an image of the food using their smartphone. These technologies primarily rely on self-recorded data by end-users. Although it is accurate, it can be tedious at times. In this work, we aim to target the advantages of utilizing speech processing, natural language processing (NLP), and text processing in nutrition monitoring. The current app in development converts speech to text using a Google's speech API. The text obtained from speech to text will be processed by NLP module for identifying nutrient information (food names, time of day, and quantity). The tagged items are then cross referenced with entries from a personal database, largely collected from the USDA's database, utilizing 2-tiers string matching algorithm, which calculates and returns their calorie values. The system has been evaluated by 10 subjects and the performance is 92.20%.

Why Using Spoken Data?

- User's hands and eyes are busy
- Keyboard is small
- User is disabled
- Spelling is important
- Faster user interaction
- Exhaustiveness of typing

Objectives

- Reinforce server security
- Collect data on speech to text accuracy
- Integrate new natural language processing API
- Report calories on app immediately after submitting
- Improve app interface and usability

Accomplishments

- Server security improvements including RSA key validation and removing root login privileges have been the solution to previous server corruption.
- Registration through app is more intuitive and auto-hides the keyboard when the page is tapped.
- Preliminary integration of API-AI to improve the natural language processing. The API also has learning features and real-time responses to prompt the user for additional information.

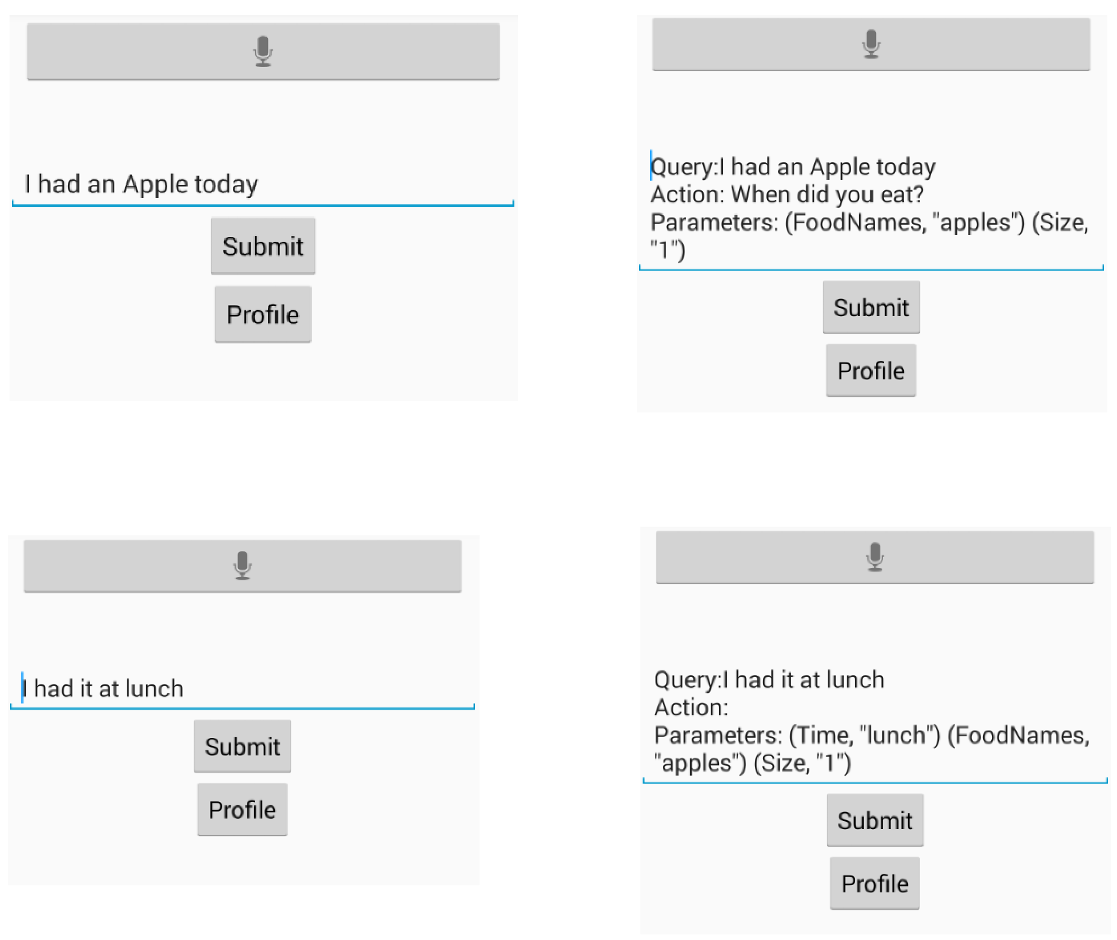

Fig. 1: Voice Recognition API-AI Inclusion

Data Collection

- 10 subjects were recruited
- A script was prepared to be read in 4 different environments (noise-free, street, music, movie)
- Participants repeated the process 3 times in each environment
- The performance of S2T (Speech2Text), NLP, and String matching modules were calculated by analyzing the data

Results

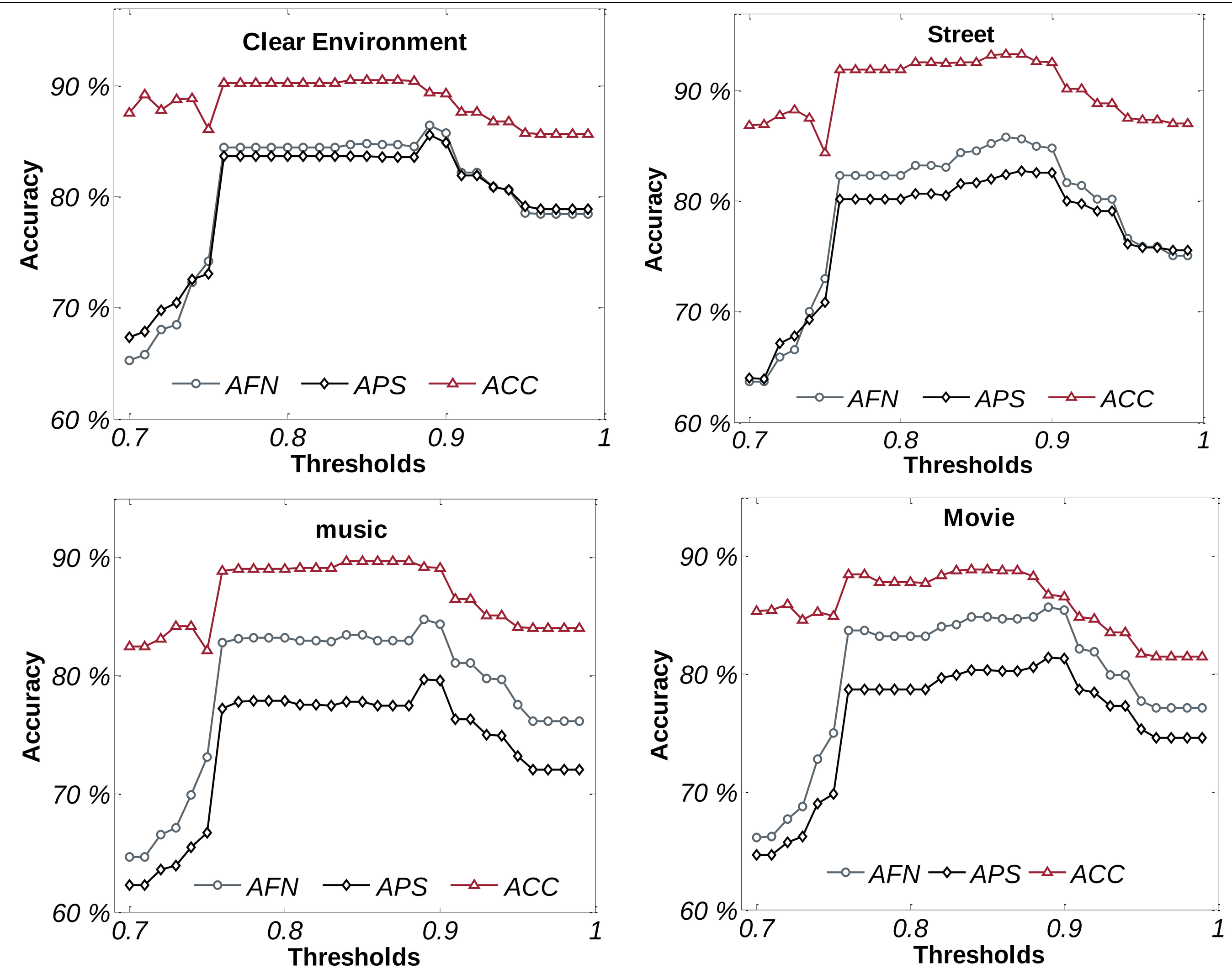

Fig. 2: Accuracy of the system utilizing NLP and string matching modules with respect to different thresholds in 4 different environments

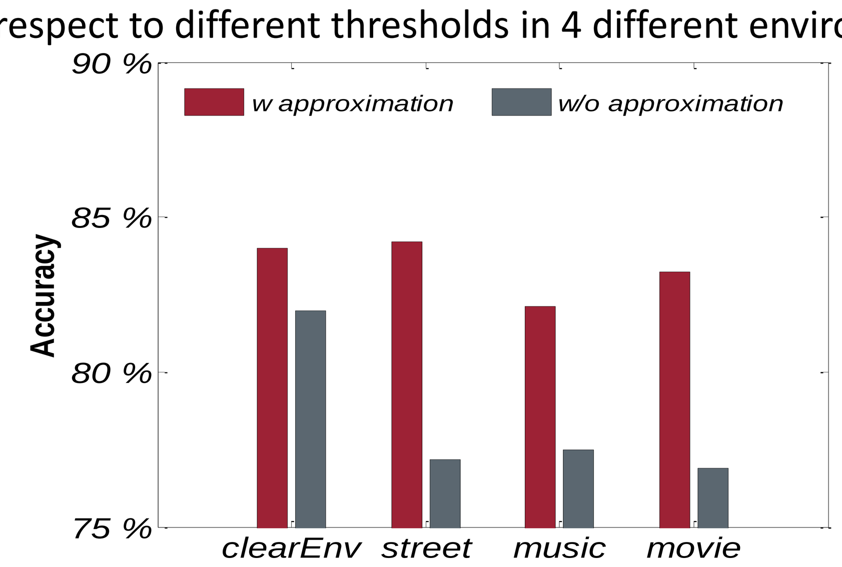

Fig. 3: Performance of finding nutrition-specific data with and without presence of approximate matching algorithm.

Conclusion

- Speech to text module is not completely accurate 81.99%
- Implementing approximation matching for increasing the performance
- Considering a similarity probability in range of 0.7 to 0.99
- System achieved the best performance around 0.85 (92.20%)

Future Work

The app currently needs a few modifications with the implementation of API-AI. The next step once those changes are made is to start usability testing. The goal is to eventually get this into the hands of a clinician who can have their clients use the app to record and monitor eating habits.

Acknowledgment

This work was supported by NIA grant R25AG046114.