Speech Coach - Final Project Report

ECE1778 - Creative Mobile Applications

Apper: Nicole Smith **Programmer:** Andrew Bitar, Ruolong Lian (Lanny)

Word Count: 1,985 + 487 (Apper Context)

I. Introduction

A. Motivation

Public speaking anxiety is the most common phobia in North American society, with approximately seventy-five percent of the population experiencing it [1]. One of the best ways to conquer this fear is to increase a speaker's confidence in their ability to deliver their message [2]. The aim of Speech Coach is to improve the confidence of the presenter, whether they are a public speaking novice or expert, by providing speakers with real-time feedback on their presentation during both practice and the presentation proper. This allows speakers to modify any "bad" behaviour that may detract from the presentation before it ever becomes obvious to the audience.

Previous studies have shown that immediate feedback increases learning compared to delayed feedback [3]. However, providing immediate feedback during a speech presents a real problem. The human working memory can only hold a finite amount of information at any time and humans generally use all of this working memory when they are producing a coherent speech [4]. Before Speech Coach, immediate feedback was difficult to provide during presentations since it required an instructor, peer, or coworker to interrupt the speaker to provide the feedback, hindering the flow of the presentation. To circumvent this issue, Speech Coach gives feedback using the phone's vibration capability, since the literature shows that providing information through different sensory modalities (i.e. visual, auditory, or tactile) decreases the cognitive load, i.e., the strain placed on the presenter's mental capacity [5]. This makes it easier for a presenter to simultaneously attend to both the tactile feedback given by the phone's vibration and the presenter's own auditory public speaking.

The stylistic features we have chosen to provide feedback for are based on the features included in most marking schemes for presentations [6] and based on the capabilities of the mobile phone. Speech Coach therefore provides quantitative feedback on a user's speech rate, their body positioning (i.e. will indicate if the user is not facing the audience), and helps users to be more self-aware about their movement during the presentation.

In particular, Speech Coach detects the rate of speech by processing the phone's microphone signal and notifies users when they exceed a certain threshold. Our default threshold is 150 words per minute (wpm) since Carver found that speeches delivered at a faster rate reduce comprehension in listeners [7]. However, we also allow users to modify this ideal speed in the settings since studies have shown that the appropriate rate of speech will differ depending on the content being delivered in the presentation [2]. For example when the content is new to the audience, speakers should slow their rate of speech [2].

Speech Coach also uses the phone's accelerometer to measure the amount of movement a speaker makes during their presentation. Some experts suggest that a speaker should walk only when transitioning between topics and this movement should be purposeful [2]. However, because of time limitations and technological constraints, Speech Coach simply alerts users when they exceed an experimentally determined movement density threshold. We plan to modify this in the future to better reflect the literature suggestions for movement.

Finally, Speech Coach gives users feedback about the direction they are facing, since speakers should face their audience during their presentation. This particular feature has a number of secondary benefits including: encouraging eye contact; discouraging reading PowerPoint slides during presentations; and increasing audience engagement [2].

B. Apper Context (written from Nicole's perspective)

Speech Coach uses some of the basic principles recommended by experts in the field of public speaking to give immediate and useful feedback to users about the quality of their presentation style. These public speaking principles are ones that I have studied extensively during my undergraduate Psychology degree at McMaster University. My experiences since then show how Speech Coach is an essential tool for:

- Students hoping to achieve better grades on their presentations
- Teachers hoping to improve the quality of their lectures
- Business sales associates who may want to refine their sales pitches

As a teaching assistant for Introductory Psychology, I often noticed that students had trouble delivering presentations. It was not necessarily a lack of content or understanding, but an absence of presentation style including avoiding eye contact, reading off of the slides, fidgeting, and other behaviours that detracted from the content of the presentation. Students would have benefited greatly from immediate feedback on their presentation. It seems unfair that the only feedback students traditionally receive comes after they have received their grade. Speech Coach rectifies this issue, as it provides feedback during both practice and the actual presentation so that students can modify their distracting behaviours and obtain better presentation grades. Speech Coach may also be a useful tool for the teacher who wishes to view the progress of the students presentation style in a quantitative way.

After my first year of teaching, I worked with a colleague to redesign a required course for first year psychology teaching assistants. A central objective of this class was having teaching assistants provide a demonstration of their lecture before they presented it formally to their students. This allowed the teaching assistants to receive feedback from other teaching assistants an invaluable process that forced teachers to self-reflect. However, professors, teaching assistants, and lecturers in general do not always have access to a community where they can receive this kind of feedback before their lecture. Speech Coach is a new way to glean the effectiveness of ones lecturing style and forces a teacher to self-reflect on their performance without needing that community.

After graduating from McMaster, I worked for a short time in marketing at Russell Investments where I was required to format and make legally compliant all of the PowerPoint presentations given by the sales associates. It was here that I realized that the same problems that troubled teachers during their lectures seem to trouble sales associates during their sales pitches. I found that the best sales associates were those who had a great delivery style rather than a thorough understanding of investments. The problem is that these individuals rarely receive any feedback on their presentation style from colleagues, mentors, or their clients. This means that self-improvement is essentially non-existent. Here, Speech Coach will also be useful, as a phones vibration feedback is not detectable by clients or other attendees, keeping the feedback discreet and professional.

II. Overall Design and Functionality

A. Front-end

The overall block diagram of our proposed application is shown in Figure 1. The red arrows represent the flow of the application from a user's perspective. The blue arrows depict the information flow among underlying modules which run in the background.

On the front-end, the application starts from the `Event Home` module (Figure 2), where users can simply swipe to the right to view the `Event Listing` module (Figure 3). From `Event Listing`, users can select a previous event from the list, or create a new event by pressing the "+" button (Figure 4). Subsequent uses of the app will automatically take the user to the `Event Home` module for their last monitored presentation.

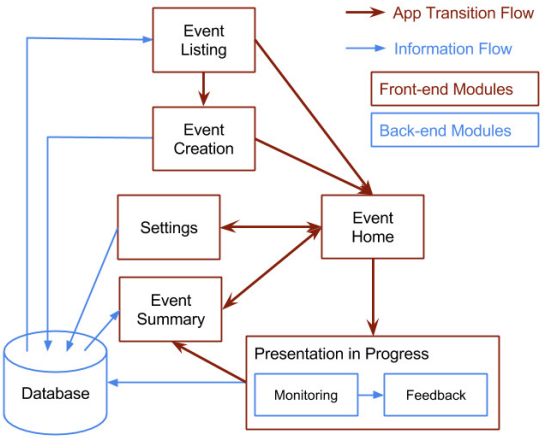

Fig. 1. Block Diagram

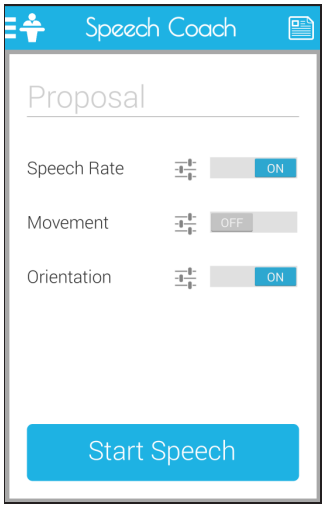

Fig. 2. Event Home

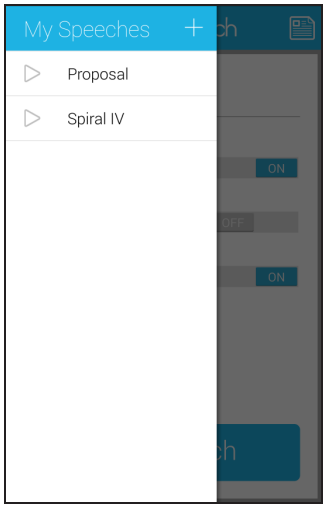

Fig. 3. Event Listing

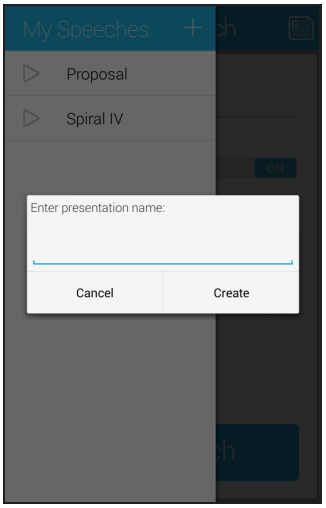

Fig. 4. Event Creation

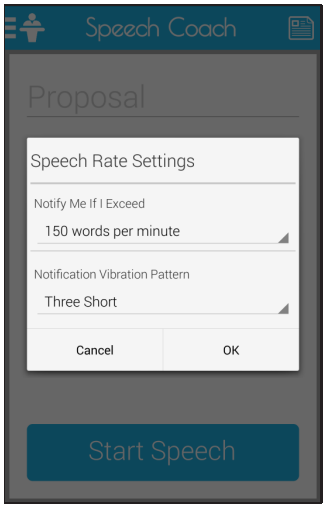

Fig. 5. Settings

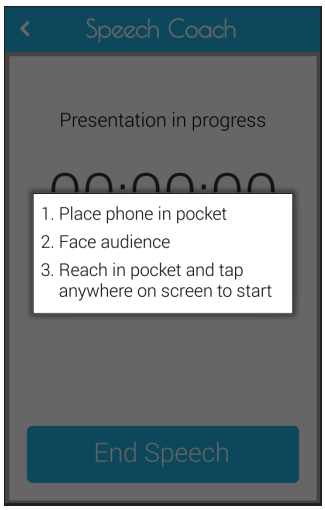

Fig. 6. Orientation Setup

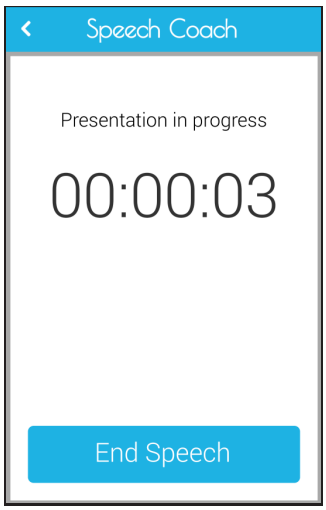

Fig. 7. Presentation in Progress

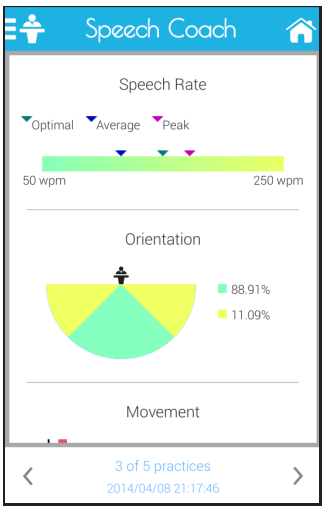

Fig. 8. Practice Summary I

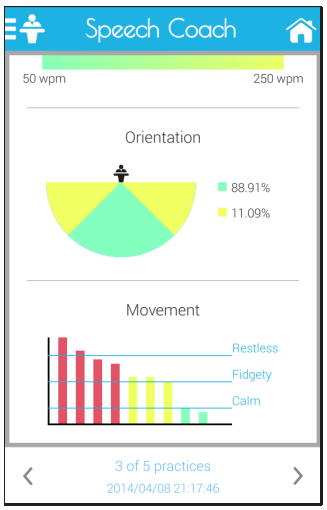

Fig. 9. Practice Summary II

Once in the `Event Home` module, users can change the `Settings` module for speech rate, movement, and orientation monitoring (Figure 5). This includes changing the type of vibration feedback associated with each monitoring option and adjusting the rate of speech. The `Event Summary` module provides summary information about past practices so that users can view their progress over time (Figure 8 & 9). Users can also begin monitoring their presentation by clicking the "Start Speech" button built into the `Event Home`. This takes users to the `Presentation in Progress` module (Figure 7). During the presentation, the back-end Monitoring module analyses the sensory input and the Feedback module provides feedback accordingly to the users. When the "End Speech" button is selected, the screen is directed to `Event Summary`.

B. Back-end

As shown in Figure 10, while the presentation is in progress, there are three key elements to the presentation that are monitored. The user's movement is detected via the phone's accelerometer. Specifically, a finely-tuned pedometer was implemented by modifying an open-source pedometer code. The code tracks the density of the user's movements. If this density exceeds an experimentally-determined threshold, the app notifies the user through the vibrator.

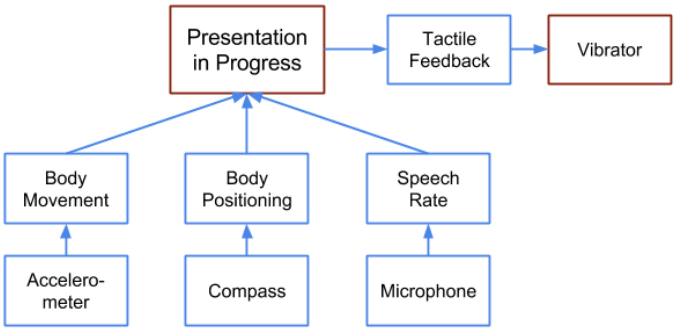

Fig. 10. Back-end Block Diagram

The direction the user is facing, i.e. body positioning, is tracked via an implementation of a compass. The compass uses accelerometer and magnetometer data to track what direction the phone is facing. When the user first begins their presentation, they set their ideal orientation. During the presentation, the app will notify the user if their current orientation is more than 45 degrees away from their ideal orientation. Poor calibration issues had to also be handled by the app, as it was discovered that the phone's magnetometer can easily become uncalibrated. To handle this, the app was modified to programmatically detect uncalibration, and to subsequently prompt the user to calibrate the device.

The speech rate is measured as the number of words per minute (wpm), which is converted from the number of detected syllables with a constant ratio of 0.66 words/syllable. The underlying syllable detection algorithm relies on the acoustic characteristic of English speech that syllables are typically vowel centric with peaks of sonority and surrounded consonants with less sonorous sounds[1]. Although the energy envelope of the audio signal is a good proxy of the sonority, it is not sufficiently accurate to detect syllables by simply counting the peaks on the energy envelope of the full band signal. This is due to the significant noise on the energy envelope of the raw signal and the fact that some closely adjacent syllables often appear as a single peak, especially in fast speech. Our implementation combines two methods proposed in [9], [10] and consists of the following major steps. The raw audio signal is first passed through four bandpass filters[2] and an energy envelope vector is obtained from each subband signal. Temporal cross-correlation is performed on each envelope vector in a shifting window fashion to emphasize a peak at every syllable's center. To construct deeper gaps between neighboring syllables, the four manipulated energy envelope vectors are then cross-correlated to form a new signal. This new signal is passed through a low-pass filter to remove small glitches. Finally, threshold mechanism examines each peak and removes the spurious ones.

[1]Sonority can be defined as a sound's "loudness relative to that of other sounds with the same length, stress, and pitch" [8]
[2]Cutoff frequencies are 300-800, 800-1500, 1500-2500 and 2500-4000 [Hz]

III. What We've Learned

Appropriate movement during presentations is difficult to define quantitatively. As previously mentioned, some experts suggest that a speaker should walk only when transitioning between topics and this movement should be purposeful [2]. However, other experts have varying opinions on this matter. Quantitatively and objectively defining good vs. bad movement has proven to be quite difficult, and will require extensive user testing. The biggest concern is ensuring that the app does not have any "false positives" - telling the user that their movement is bad when, in fact, it is perfectly acceptable.

Initially, the plan was to rely on the phone's vibrator to provide real-time feedback on a user's presentation. Though there was some concern on the user's ability to understand the feedback, we believed that having very distinct vibration patterns would facilitate this. Based on our experience using the app, we have learned that a user will likely be so focused on their presentation that they will not be able to distinguish different vibration patterns. The most that can be realistically expected from the user is the ability to distinguish whether their phone is vibrating or not. As a result, we learned that visual feedback will ultimately be necessary for the app to provide different kinds of feedback to the user in real-time.

The immense complexity of implementing an accurate speech rate detection algorithm quickly became evident during the development of Speech Coach. It was very challenging to fine-tune the algorithm such that the speech rate can be accurately detected for different users. We learned that it is essential to set up an experiment framework with a comprehensive set of audio testbenches, in order to allow efficient evaluation of our filter design and parameter settings.

IV. Contribution by Group Members

Nicole contributed to the project by advising the programmers based on research in psychology and education. Specifically, she researched the appropriate thresholds and quantitatively defined good and bad habits during speeches. Nicole was also in contact with Melanie Novis who provided guidance for defining 'good' movement during presentations. Another part of Nicole's contribution was designing the summary page feedback since Hattie and Timperley found that if feedback is not provided well, it can be detrimental to learners [11]. Part of this process was utilizing data visualization and ensuring that feedback was provided clearly and concisely so that users are not overloaded with information. Nicole also wrote and designed the first drafts of all presentations and the written assignments.

Both Andrew and Lanny worked collaboratively on the app's UI/UX design. For the back-end, Andrew was responsible for implementing body movement and orientation detection, while Lanny was responsible for implementing speech rate detection.

V. Future Work

There is still much fine-tuning that could be done to increase the proficiency of speech rate detection. Additionally, we hope to better quantitatively define good and bad movement during presentations. We also want to expand the capabilities of the application to provide a recording option for playback as Melanie Novis recommended. This would allow users to evaluate for themselves the strengths and weaknesses of their presentation and may provide a more qualitative assessment of the presentation [2]. For example, posture is difficult to measure using a mobile phone application, but should be easy to detect if users are reviewing their performance on camera.

We also hope to synchronize the phone to the laptop so that users can pace themselves according to the slide deck. Ideally, the laptop will display green if the presenter is on-pace, yellow if they are taking slightly longer than

they planned or practised (i.e. less than 5 minutes over time), and red if they are over their ideal time by more than 5 minutes. We feel that this type of feedback would be useful to novices and experts alike. Visual feedback will also help us overcome the limitations of the vibrator feedback discussed earlier. We have also considered adding additional features including feedback on the user's volume, and/or a count of word repetition ('um'-s, 'so'-s, etc.).

REFERENCES

[1] "Fear of public speaking statistics." http://www.statisticbrain.com/fear-of-public-speaking-statistics/, Nov. 2013.

[2] M. Novis, *Canadian Public Speaking*. Pearson Education Canada, 2003.

[3] J. Kulik and C. Kulik, "Timing of feedback and verbal learning," *Education and Educational Research*, vol. 58(1), pp. 79–97, 1988.

[4] J. Kim, *Introduction to Psychology Handbook*. Pearson Education Canada, 2011.

[5] R. E. Mayer and R. Moreno, "Nine ways to reduce cognitive load in multimedia learning," *Educational Psychologist*, vol. 38(1), pp. 43–52, 2003.

[6] J. S. Anderson and L. C. Mohrweis, "Using rubrics to assess accounting students writing, oral presentations, and ethics skills," *American Journal of Business Education*, vol. I(2), pp. 85–94, 2008.

[7] R. P. Carver, "Effects of increasing the rate of speech presentation upon comprehension," *Journal of Educational Psychology*, vol. 65(1), pp. 118–126, 1973.

[8] I. P. Association, *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge University Press, 1999.

[9] D. Wang and S. Narayanan, "Robust speech rate estimation for spontaneous speech," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, pp. 2190–2201, Nov 2007.

[10] N. Morgan and E. Fosler-Lussier, "Combining multiple estimators of speaking rate," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 2, pp. 729–732 vol.2, May 1998.

[11] J. Hattie and H. . Timperley, "The power of feedback," *Review of Educational Research*, vol. 77(1), pp. 81–112, 2007.