

Correctly: Android App to Help Pronounce Better

Venis Nasr^{*}, Nishant Arora⁺, Salavat Nabiev^φ

^{}Ontario Institute for Studies in Education, University of Toronto, Canada*

⁺^φDepartment of Computer Science, University of Toronto, Canada

E-Mail: ^{}venis.nasr@mail.utoronto.ca, ⁺nishant@cs.toronto.edu, ^φsalavat@cs.toronto.edu*

Project: <https://github.com/whizzkid/correctly>

Abstract - We propose an android app which helps users to acquire a new language quickly and accurately. This app focuses on the speaking aspect of language acquisition. It helps users to learn how to pronounce new words in order to convey their messages effectively. This facilitates in socialization and help them perform their daily tasks. Furthermore, we intend to utilize this technology to help teachers working with learners, and learners working independently, to take valuable steps that makes their language development possible. The app helps users to record their voices match their pronunciation and provide feedbacks whether their pronunciation is correct or not. We call this app ‘Correctly’ and is referred by that name in the following text.

I. INTRODUCTION

Canada is the first country to adopt multiculturalism as an official policy. Millions of newcomers land in Canada every year, coming from different countries and speaking different languages. Acquiring a new language is based on developing four skills reading, writing, speaking and listening. The Common European Framework of Reference for Languages: underlines doing things with language rather than just learning about language. Many of different aspects of language can be practiced with the use of self-study materials. They can do as many grammar exercises as they want; they can work with audio and video to improve listening; however it is not likely these type of exercise will correct the students’ pronunciation. Pronunciation is one of those things that only teachers in classroom can correct; it’s a labor intensive aspect of language acquisition.

II. STATEMENT OF FUNCTIONALITY

Users will start by learning new words. Users are expected to practice the meaning and pronunciation of each word on the wordlists built in the app. Forcing language learners to rush into sentence formation can interfere with vocabulary learning during the early stages of acquiring a new language. Instead, learners should be given time to acquire the meanings and pronunciation of individual words at their own pace before being required to use them in a larger context ^[2]. Correctly uses a built in word lists that includes the most frequent used words. Language learners who take that time are far more likely to use and pronounce the words correctly when they do choose to form sentences ^[2].

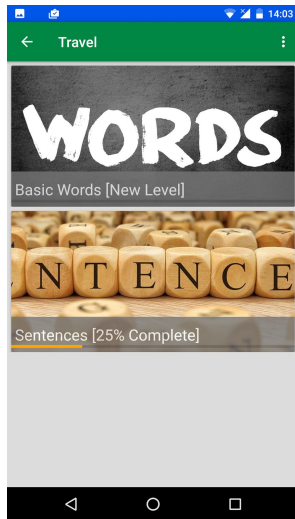
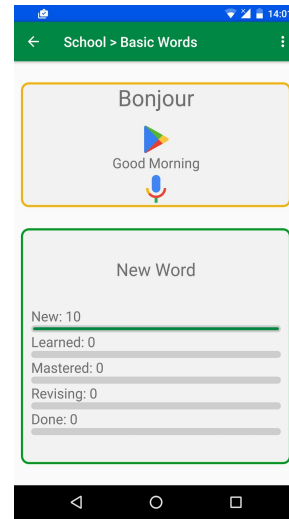
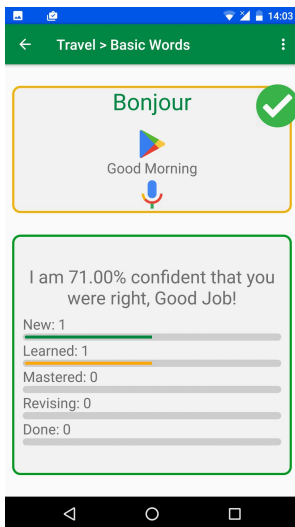


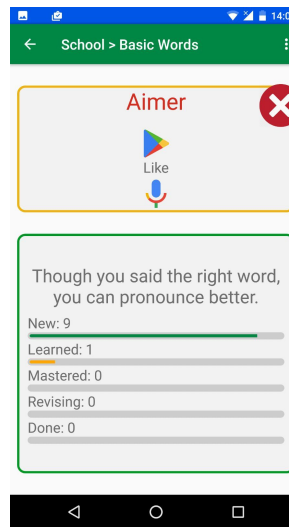
Figure 1: users start by selecting words.



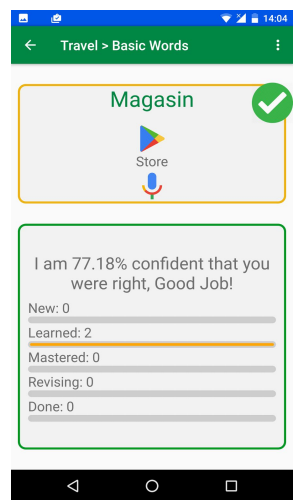
Users can listen to the correct pronunciation



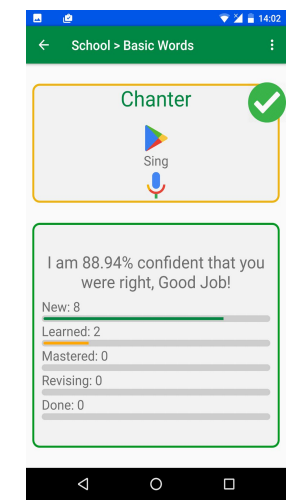
words highlighted green if pronounced correctly



Words highlighted red if pronounced wrong

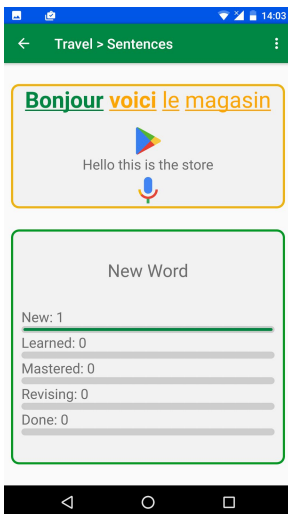


Progress indicators display total progress

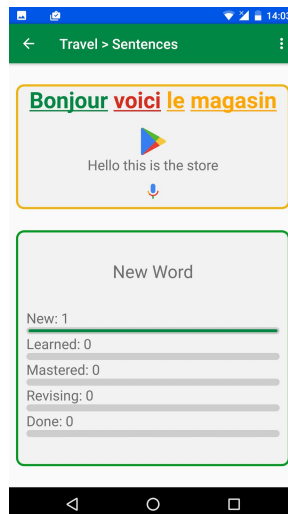


Progress is relative to all words in store

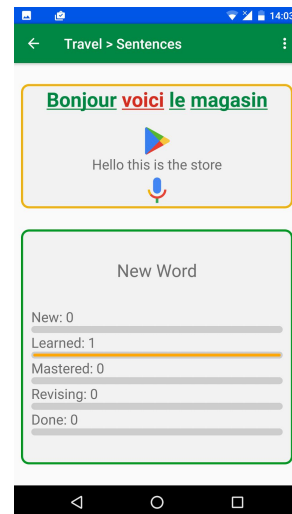
Second phase is the sentence phase; users unlock the sentence phase by completing all words in the word phase. Users will practice reading sentences that includes the words acquired in the first phase. The more frequently language learners are exposed to foreign vocabulary; the more likely they are to remember it. Studies suggest that most learners need between 5-16 'meetings' with a word in order to retain it^[2]. When the word has to pass by four different levels before unlocking the sentence phase and the number of times user meet the word in the sentence phase, Correctly guarantees a minimum of 7 meetings with the same word. Language learners are thus more likely to use and enjoy the app long enough to accomplish a sufficient number of 'meetings' to master the new vocabulary terms^[2].



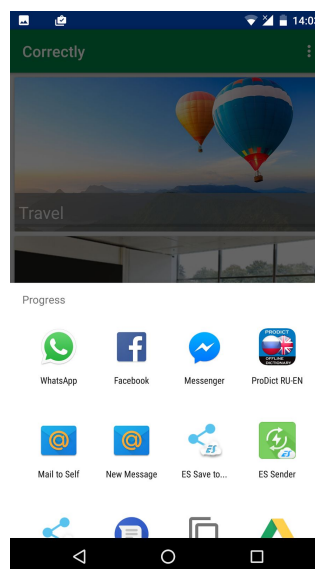
Users can listen the correct pronunciation of the full sentence or individual words



Users pronounce each word of the sentence and Correctly provides instant feedbacks by highlighting the word red if mispronounced and green if pronounced correctly.



We accept the user's pronunciation as correct if it has a single error because contextually it will make sense.

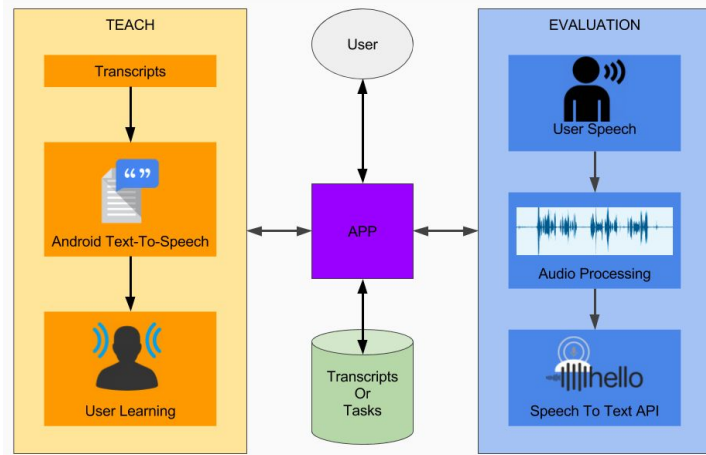


Users can share their progress

The team had an idea for an interesting functionality but due to time constrain, the team couldn't develop it. The live mode is intended to help users practice their pronunciation freely. Users can record any speech and the app will detect mispronounced words based on language structure and words context.

III. OVERALL DESIGN

We intend to utilize the Google Speech Api^[1] to help with speech recognition. This task is more effectively visualized in the following block diagram:



Google's speech API^[1] is still in beta and can help recognise over 80 languages. Here is a short snippet demonstrating how we communicate with the API:

```
{
  "config": {
    "encoding": "FLAC",
    "sample_rate": 16000,
    "language_code": "en-US"
  },
  "audio": {
    "uri": "gs://speech-demo/shwazil_hoful.flac"
  }
}
```

Sample gRPC request to Google Speech API^[1]

```
{
  'results': [
    {
      'alternatives': [
        {
          'confidence': 0.98267895,
          'transcript': 'how old is the Brooklyn Bridge'
        }
      ]
    }
  ]
}
```

Sample Response from Google Speech API^[1]

The Google Speech API^[1] returns us with a transcript detected in the audio and a “Confidence Value”^[3]. Understanding this confidence score is the key to evaluating user performance. As per the documentation, the confidence score is defined as follows:

“When returning an alternative, the Speech API will assign a confidence value to any given transcription, on a scale of 0.0 to 1.0, with 1.0 meaning absolute confidence. You can use these confidence values to compare alternatives, or to decide whether to return results to a user (and/or ask for confirmation from the user)”.

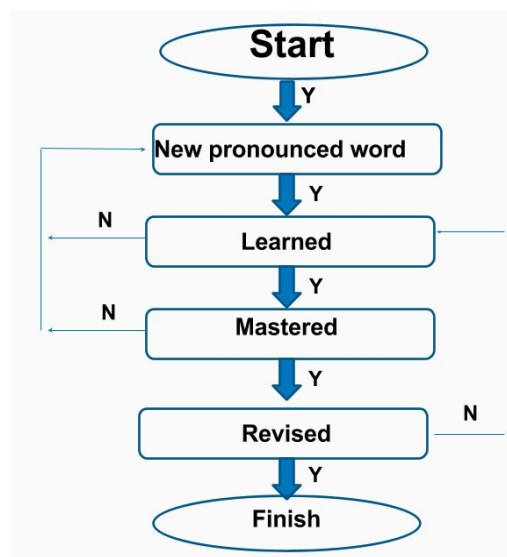
This essentially means, that this is a probability assigned to generated transcript which implies we can make use of these values to evaluate user's speech. Since the speech api focuses on returning alternatives we can make the following confusion matrix:

Parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Audio Spoken	America	Table	Chair	Car
Transcript	America	Table	Hair	Cart
Confidence	0.9	0.4	0.3	0.7
Alternatives	None	Cable, Fable	Care, Fare	Bar, Far
Evaluation	Good	Good if confidence > 0.65 else Bad	Bad	Bad

So we encounter four scenarios returned from Google Speech API:

- 1. Transcript Matches With High Confidence:** In this case, we can safely assume that the user spoke the right word and we can evaluate this to Good.
- 2. Transcript Matches With Low Confidence:** This is a pretty common scenario, in this case, the word sounds right, but the speech API is not confident enough. This is most likely because there are a lot of similar sounding words in that language. Which implies that we can safely assume this word to be correct if it's above a certain threshold. A research was performed to come up with an average to the least accepted confidence level to consider the word pronounced correctly.
- 3. Transcript Mismatches With Low Confidence:** This is likely a wrong pronunciation and we report this as bad.
- 4. Transcript Mismatches With High Confidence:** This is definitely a mis-pronunciation and we report this as wrong.

Once we have made sure that we can safely evaluate pronunciation for every instance, we now need to make sure that the user has enough interactions with the word so that they are able to retain it. Research suggest that most learners need between 5-16 'meetings' with a word in order to retain it^[1], we devised the following word mastering algorithm to solve this:



This algorithm always makes sure that the user gets to see every word at least five times before we mark it as complete.

IV. REFLECTIONS

Group dynamic and team work is one of the most important lessons to reflect on. Group dynamic is the behavioral relationships between members of a group that are assigned a common task. These dynamics are affected by roles and responsibilities and have a direct influence on the product. According to Bruce Tuckman^[4], group formation encounters several developmental phases with similar conflicts and resolutions. We analysed the following:

- In the forming stage^[5], the team should meet and learn about the opportunities and challenges, and then agree on goals and begins to tackle the tasks.
 - Correctly team didn't meet frequently enough.
 - Team members tended to behave quite independently.
 - It was essential at this phase that members attempt to become not only oriented to the tasks but to one another as well.
- The second phase is the storming phase^[5], this phase is focusing on the personalities and how team members are interacting with each other. Team members need to emphasize their differences and to make it clear that without tolerance and patience the team will fail.
 - Correctly team didn't benefit from the disagreements within the team in order to make members stronger, more flexible, and able to work more effectively as a team.
 - The result was evident in the spiral 2 presentation.
- Moving on to the third phase or the norming phase^[5]. in this stage, all team members take the responsibility to work for the success of the team's goals.
 - We started accepting each other as they are and we made an effort to move on.
 - This was evident in the development of the app and how spiral 4 presentation was successful when compared to spiral 2 one.
 - However, the team was so focused on preventing conflict that we were reluctant to share controversial ideas; especially that team members are coming from different disciplines.
 - This point is essential to be aware of and to try to avoid it as it might not be as effective for the project in general.

As a team, we learned how to explain intricate knowledge to each other effectively, although we're coming from different disciplines and different background. This point specifically is so crucial; to benefit the community and society. Researchers from different fields of work, studies and researches can cooperate if communication is clear and open.

V. CONTRIBUTIONS

Venis Nasr MED. student (Specialist)

- Provide the French materials appropriate to develop the app.
- Perform linguistic research to come up with acceptable confidence level.
- Schedule all meetings and generate meeting agenda.

- Determine, communicate, and track meeting action items.
- Monitor project timelines, identify milestones, and monitor deadlines.
- Develop all presentations and documents submitted.

Nishant Arora M.Sc. Applied Computing (Programmer)

- Ideated about using Google Speech API^[1] for this task.
- Implemented gRPC^[6] and ProtoBuf^[7] based streaming clients interaction with Google Speech API^[1].
- Designed and Implemented data store based on Realm to store challenges, levels and progress.
- Designed and Implemented word mastering algorithm.
- Handled most of the backend tasks.
- Wrote animated fragment loaders for easy transition management between fragments.
- Peer reviewed contributions.

Salavat Nabiev M.Sc. Applied Computing (Programmer)

- Worked primarily on the frontend.
- Developed and implemented UI (design elements, animations, structure, etc.)
- Implemented Text to Speech pronunciation of the words and sentences.
- Implemented word highlighting mechanism, based on confidence and transcript extracted from speech API responses.
- Handled some of the backend tasks.
- Implemented word to word transition in the sentences level.
- Peer reviewed contributions.

VI. SPECIALIST CONTEXT

Many people are trying to acquire new language for a reason or another. It depends on their native language or their mother tongue; learning a new language might be really easy or it may require more effort and time. Second language or the targeted language may have some sounds that the native language might not have, so language learner will have to learn how to make completely new sounds. Accordingly, it becomes very hard to reproduce the correct pronunciation. In order to produce any sound, we use facial musculature which includes everything from the jaw and the tongue. There are also other elements to consider when speaking such as where the tongue touches the teeth or which palate speaker is employing when shaping vowels in conjunction with the tongue.

Correctly app can be used to perform some researches in order to facilitate the pronunciation aspect of the language. I can gather information such as user's mother tongue and perform some research on this language, for instance, the muscular parts of the face used to produce certain sounds. Then, monitor the progress of the user and record the hardest sounds to produce. Finally, compare these sounds with the sounds produced within the mother tongue; to come up with recommendations for language learners on how to benefit from their mother tongues and use it as an asset to improve the targeted language or at least what difficulties prevent them from producing the correct sound. For instance, when Arabic speakers learn English, they pronounce words with P in English, like a B, they have difficulty differentiating between 'p' and 'b'; since Arabic has no 'p' sound speakers will often say 'p' as 'b' like Bolice instead of Police. Another example for English speakers who are learning French, when they

pronounce the letter U it's really hard and it's completely different than the sound in English; the lips are saying "oo" as in book while the tongue is saying "ee" as in free.

Correctly can provide data to discover such pattern between different languages and provide recommendations to teachers and language professors to help language learners overcome such barriers. The app might provide a lot of data for studies regarding English Language learners who have unusual mother tongue.

VII. FUTURE WORK

The app can be developed from different direction in the future. First, the app highlights the mispronounced syllable of the word; this can help more advanced users to fix their small mistakes. Second, teachers has the flexibility to add their own words in order to plan and implement their own lessons, that best suits their students' needs. Also, the app can be further developed to take picture of a certain book, convert it into PDF then convert the text into speech. The user can listen to the correct pronunciation, then integrate the previous functionality with Correctly to check their own pronunciation. The implementation of the Live mode when users can record any speech and the app will detect mispronounced words based on language structure and words context. The app can provide tips to the user on how to best pronounce certain pattern For instance, nasal sounds in French (an, on) or the R, based on the user's mother tongue.

VIII. SUBMISSIONS

<i>Items</i>	<i>Post on the course Website</i>
<i>Video of Final Presentation</i>	<i>ok</i>
<i>Report</i>	<i>ok</i>
<i>Source Code</i>	<i>ok</i>

IX. REFERENCES

- [1] "Speech API - speech recognition | Google cloud platform," Google Cloud Platform. [Online]. Available: <https://cloud.google.com/speech/>. Accessed: Dec. 14, 2016.
- [2] J. Barcroft, "Second language vocabulary acquisition: A lexical input processing approach," *Foreign Language Annals*, vol. 37, no. 2, pp. 200–208, May 2004.
- [3] "Cloud speech API basics," Google Cloud Platform, 2016. [Online]. Available: <https://cloud.google.com/speech/docs/basics#confidence-values>. Accessed: Dec. 14, 2016.
- [4] Tuckman, B. W, Developmental sequence in small groups. *Psychological Bulletin*, vol. 63, no.6, pp. 384-399, 1965.
- [5] Tuckman, B. W, Developmental sequence in small groups. *Research and Application Journal* pp. 71-72, 2001.

- [6] "Package google.rpc | Google cloud speech API (limited preview) | Google cloud platform," Google Cloud Platform, 2016. [Online]. Available: <https://cloud.google.com/speech/reference/rpc/google.rpc>. Accessed: Dec. 14, 2016.
- [7] "Protocol Buffers | Google developers," Google Developers. [Online]. Available: <https://developers.google.com/protocol-buffers/>. Accessed: Dec. 14, 2016.