CS224S Assignment 1: Speech Systems and Phonetics

Borislav Mavrin

Winter 2021

1 [30 pts] Speech APIs and Personal Assistants

1.1 Speech Transcription

1. Simple messages results from iOS email dictation:

   hey how are you I was thinking if you prefer you can come to our dinner party tomorrow please let me know if you're coming or not

   *hi I'm planning to go for a vacation probably in Mexico let me know if you want to join us in **the strip** and I will book tickets for you*

   The only error is **the strip**, which has to be **this trip**.

2. The system does not guess the punctuation, but allows punctuation commands, e.g. *"comma", "period", etc.*

3. As it can be seen from the following example: *hey I'm planning to leave at five delete that I meant six* the system does not support corrections. However, it does correct words on-the-fly given more context.

4. I was able to break the system by using domain specific terminology as in the following example:

   original: *GANs or generative adversarial networks in machine learning. This is a very cool field of AI.*

   result: *Gantz or generator for the referral networks in machine learning this is a very cool field of 38*

   The system, apparently, does not have words like *"GANs", "adversarial"* in its dictionary or does not handle them well.

   I also tried switching from English to Russian within single utterance:
   original: *Hey how are you? ¡continued in Russian¿*

   result: *hey how are you created*

   The user have to switch languages explicitly.

1.2 Personal Assistants

I used Google Assistant on Google Home for the experiments in this section.

1. The system answers factual questions accurately and supports follow up questions. The answers are given in complete sentences.

   me: *"Who wrote The Great Gatsby?"*

   G: *"F. Scott Fitzgerald wrote The Great Gatsby."*

   me: *"When was that book published?"*

   G: *"The Great Gatsby was first published on April 10, 1925."*

2. I asked the system for burger places near me. The system suggested one place and provided the rating in the same interaction, I tried asking about the menu - it gave me more locations of the same chain. I tried a few times but couldn't get past two interactions. The interaction was completely speech driven.

3. The system failed to process a lengthy initial command with time, attendees and location. However, when I started with a simple "create an event", the system asked me about the time and title of the event. I followed up by adding location.

4. I tried a few themes, but the system didn't allow me to barge-in without wake word. Then I tried to barge in with a wake word and the system allowed me to make corrections.

   me: *"Hey Google. Who is stronger Thor or Batman?"*

   G: *[long response]*

   me [during response]: *"Hey Google. I meant Hulk."*

   G: *[gave a response to "Who is stronger Thor or Hulk?"]*

5. All the issues I had can be contributed to the limited functionality and issues with speech recognition.

1.3 Text-to-Speech (TTS) API

1. I was able to understand all words easily. But the voice does not sound human. I noticed abrupt changes in tone, noise, lack of pauses and no changes in intonation.

2. The system tries to change the intonation in sentences with "!" and "?". But it does not feel natural.

3. Adding pauses does help - but I still hear sharp noises.

2 [20 pts] Phonetic Transcription

2.1 Correct mistakes

We often process speech data in phonemes instead of words. Find and correct the mistakes in the ARPAbet transcriptions of the following words:

1. three [dh r i] [th r iy]
2. sing [s ih n g] [s ih ng]
3. eyes [ay s] [ay z]
4. study [s t uh d i] [s t ah1 d iy]
5. though [th ow] [dh ow]
6. planning [p pl aa n ih ng] [p l ae n ih ng]
7. slight [s l iy t] [s l ay t]
8. action [ae k t ah n] [ae k sh ah n]
9. tangle [t ae ng g l] [t ae ng g ah l]
10. higher [hh ay g er] [hh ay er]

2.2 Transcribe

Transcribe the following words into ARPAbet.

1. red [r eh d]
2. blue [b l uw]
3. green [g r iy n]
4. yellow [y eh l ow]
5. black [b l ae k]
6. purple [p er p ah l]
7. suit [s uw t]
8. wash [w aa sh]
9. water [w ao t er]
10. watermelon [w ao t er m eh l ah n]

NOTE: I used CMU's resource at `http://svn.code.sf.net/p/cmusphinx/code/trunk/cmudict/cmudict-0.7b`

3 [50 pts] Getting Comfortable with Speech Data

See the Assignment 1 page on the website for submission instructions.