

Report on Lab 3

I first selected a domain: *Coffee*. My final project is related to *order* occasion, and the coffee area exactly has a lot of specific terms about the origin, category, brewing method, etc.

So I built up my plain text with 5 sentences like this:

Common coffee brewing equipment includes Chemex, Syphon, Aeropress, French Press, and Moka Maker.

Common coffee classifications include Decaffeinated Coffee, Half-decaf coffee, Fresh Ground Coffee, Black Coffee, and Instant Coffee.

Cafe Coffee: Espresso, Americano, Cappuccino, Mocha, Caffè Latte, Macchiato, Pour Over Coffee, Dutch Coffee or Cold Drip Coffee, Ice Coffee, Cold Brew, French Coffee, Drip Bag Coffee

Three major types of coffee: Arabica, Robusta, Liberica

Common coffee categories: Bourbon, Caturra, Catuai, Typica, Geisha

Then I tried three methods to do the transcription:

Firstly, I tried to read the content by myself. I realized that I need to read slowly word by word or it will easily fail to recognize the correct content. And I even need to check the pronunciation in advance for some words are not so common.

Later, I tried to use *Google Translation* to generate the audio, and I noticed that its default setting is UK accent, which tended to transcribe "Syphon" as "Siphon" (two different word forms for the same word).

When I used *Instant speech to text* to test my guess, I found there literally existed a few words that could not be recognized correctly. For example, "Chemex" (one kind of American filter pot) was always recognized as "Chermex" and "Catuai" (a kind of coffee bean variety) was always transcribed as "KATU I".

For test more occasions, I tried with my name and my friends'. It is easy to find that the pronunciation of lots of Chinese characters is not so understandable. For example, my name "Xiumei" was always transcribed as "show me". The characters composed by easy syllables like "Lin", "Chao", and "Zhao" have a greater chance of being identified correctly.

From the findings above, I trained and deployed my model at last. But when I want to download the log files, Custom Speech interface always showed "No messages and transcript records were returned for the endpoint xx ". It seemed that I didn't call my model successfully though I had completed the conversation with my new dmMachine. It could be a problem. I tested online and it showed that WER is around 13.70%. "Chemex" has been transcribed successfully but "Catuai" still not.