Lab 4A Report

Test cases

Words/Phrases	Confidence Score
Batman	0.061666097 (**LOW**)
Pikachu	0.8425556 (**HIGH**)
Angkor Wat	Unrecognizable
Supercalifragilisticexpialidocious	No reaction
The word is supercalifragilisticexpialidocious	0.551779 (**MEDIUM**)

Why Batman Struggled

Despite being a famous superhero icon, Batman suffers from phonetic ambiguity. It sounds similar to common phrases like "bad man" or "that man". In a generic language model, these common phrases would typically have a much higher probability than Batman. Therefore even if the ASR detects the sounds correctly, the confidence score remains extremely low because the model biases toward the more statistically likely phrase.

Why Pikachu Succeeded

Although it is also a popular pop-culture icon, Pikachu is the opposite of Batman in terms of its phonetic properties. It does not sound like any other common English words, making it easily recognizable. Additionally, it is likely included in the base language model due to its global popularity.

Why Angkor Wat Was Unrecognizable

This is a classic Out-of-Vocabulary (OOV) situation. Base ASR models are trained on broad and generic datasets. Angkor Wat, a temple in Cambodia, is a niche term that is highly unlikely to be in the model's dictionary, Thus the ASR attempts to match the sounds to something unrelated.

Why Supercalifragilisticexpialidocious Has No Reaction

It seems that extremely long words like Supercalifragilisticexpialidocious resulted in no reaction from the ASR at all. This is an edge case where a very long word is interpreted as a long silence, or there is some sort of buffer threshold that it has exceeded.

But The word is... works

Interestingly, the ASR is able to capture the input when the long word is spoken together with context. This demonstrates how linguistic context can impact recognition accuracy.

Solution (VG)

This demonstrates that the base ASR model is inconsistent with cultural terms and fails when faced with niche vocabulary.

A solution for such an issue is to use Azure Custom Speech (VG part). By providing a vocabulary list containing these specific terms, we can train a model that will recognize these as valid entities. This significantly improves the confidence score of terms like Batman and makes Angkor Wat recognizable.

However, the edge case for long words like Supercalifragilisticexpialidocious is still not resolved in isolation. This means that configuration tweaking (such as adjusting silence timeouts in the SDK) or encouraging users to provide context might be the way to resolve this.

Endpoint ID: 62b8c588-0191-4fd7-a7cc-1b437963d69b

Outcome (VG)

Words/Phrases	Confidence Score
Batman	0.7923598 (**HIGH**)
Pikachu	0.88747746 (**HIGH**)
Angkor Wat	0.7247133 (**HIGH**)
Supercalifragilisticexpialidocious	No reaction
The word is supercalifragilisticexpialidocious	0.95502687 (**HIGH**)