Influence Consulting

08/18/2023

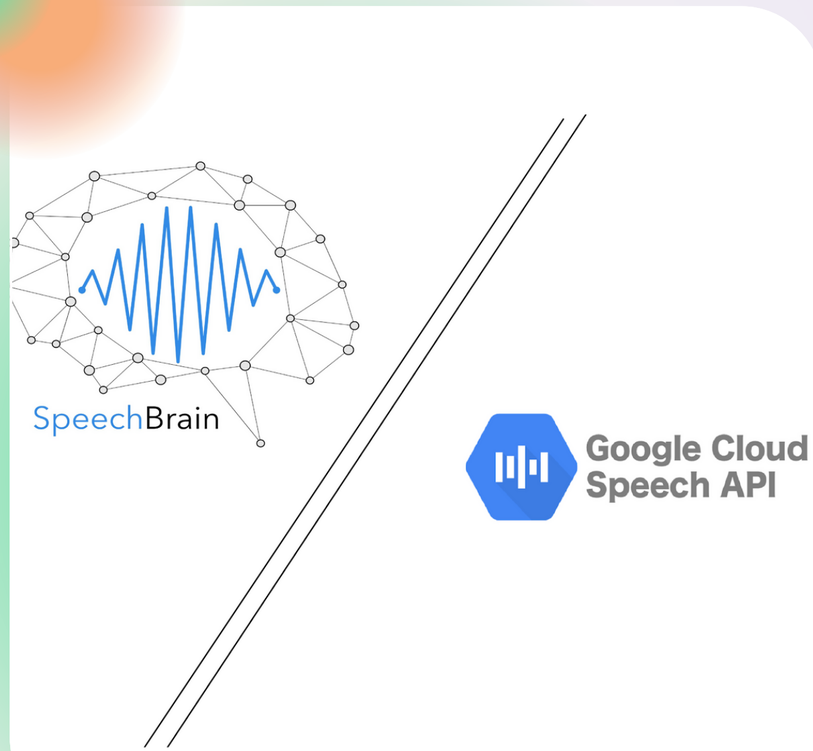
PRESENTED BY

Bradai Omar
Karoui Amine

Automatic speech recognition system for Tunisian dialect

Add a short description here

Purpose of the Presentation

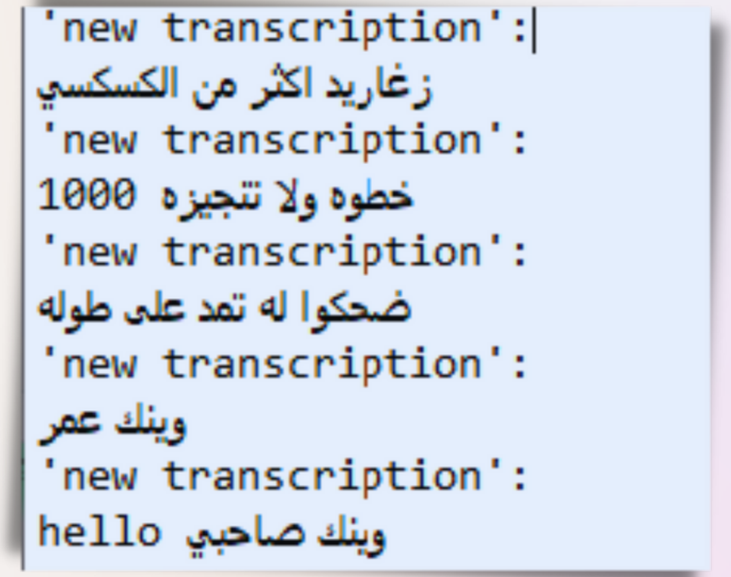


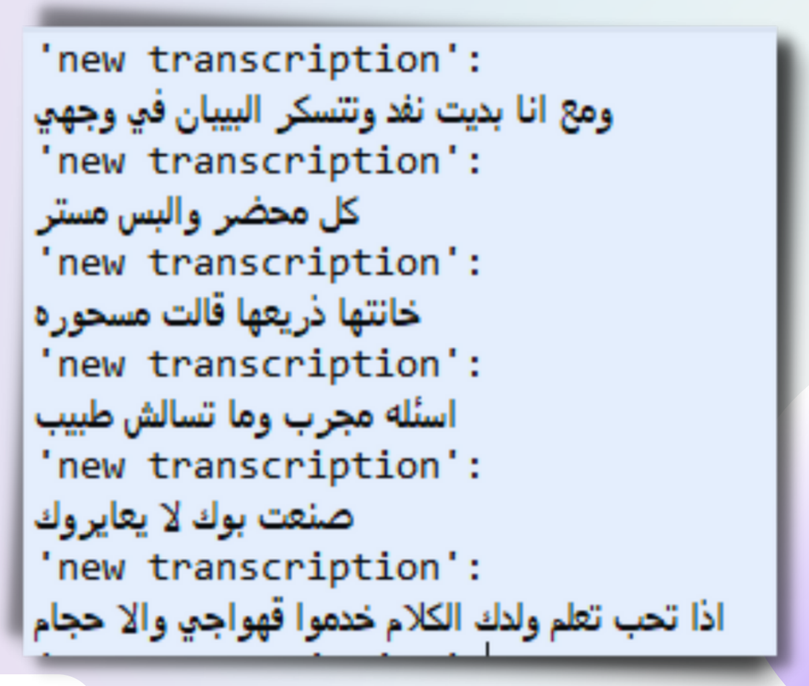
The primary objective of this presentation is to conduct an in-depth comparison of two Automatic Speech Recognition (ASR) solutions that are designed specifically for the Tunisian Arabic dialect. By undertaking this comparison, we aim to provide a comprehensive understanding of the strengths and weaknesses of each solution, their precision levels, and other pertinent factors.

This analysis will enable us to make informed decisions about which solution might be more suitable for various applications and scenarios.

Solution1: Overview

The Google API for the Tunisian dialect is a powerful tool that leverages Google's robust Automatic Speech Recognition (ASR) technology to transcribe spoken Tunisian Arabic into text. This technology is built upon Google's extensive experience in speech recognition and language processing, making it a widely recognized and accessible solution.





Examples of transcribed tunisian phrases through a microphone input

Pros and drawbacks of the Google API

The Google API (Pros)	The Google API (Cons)
• **Ease of Use:** Integrating with Google APIs is generally straightforward, thanks to well-documented resources, code examples, and developer support. This means faster implementation for projects. • **Language Support:** While not tailored specifically for the Tunisian dialect, Google's ASR technology has proven to be effective in recognizing a wide range of languages and dialects. This versatility could contribute to accurate transcription of Tunisian & Arabic.	• Limited Customization: The free version of the Google API might not provide extensive customization options that are optimized for the nuances of the Tunisian dialect. Customization is essential to accurately capture local vocabulary, pronunciation, and speech patterns. • Privacy and Data Concerns: Using a third-party API involves sharing audio data with Google's servers, which could raise concerns about data privacy and security, especially for sensitive or confidential content.

Pros and drawbacks of the Google API

The Google API (Pros)	The Google API (Cons)
• **Performance:** The precision of Google's ASR technology is generally high due to its vast training data and advanced algorithms. It can offer competitive accuracy when transcribing speech, including Tunisian Arabic**. (an average of 30% WER)** • **Potential for identifying additional languages within the sentence, which proves advantageous given the frequent incorporation of multiple languages within the Tunisian dialect.**	• **Reduced Accuracy for Specific Dialects:** Since the service is not specifically designed for regional dialects like Tunisian Arabic, it might struggle to accurately capture unique pronunciations, vocabulary, and speech patterns, leading to potentially less precise transcriptions. (DUE TO THE FREE VERSION) • **Dependency on Internet Connectivity:** The free version operates online, relying on a stable internet connection. This can be limiting in situations where offline transcription or reliable connectivity is necessary. • **Usage Limits:** The free tier comes with limitations on the amount of audio data that can be processed within a given time frame.

Solution 2 : Overview

This solution involves a customized ASR system developed using BrainSpeech technology, coupled with a Tunisian language model.

We used BrainSpeech. This technology encompasses advanced algorithms, models, and methodologies that enable improved accuracy and performance in speech recognition tasks


```
speechbrain.core - Beginning experiment!
speechbrain.core - Experiment folder: partly_frozen_splitted_wavlm/1986/
['<blank>', '<bos>', '<eos>', '|', 'ف', 'ر', 'ى', 'ن', 'ب', 'ح', 'ل', 'ه', 'ك', 'ي',
'و', 'م', 'آ', 'ؤ', 'ء', 'ئ', 'ث', 'ع', 'ذ', 'ز', 'ظ', 'ج', 'خ', 'ط', 'ص', 'ض',
41
Loading the LM will be faster if you build a binary file.
Reading C:\Bradai\focus\AI ASR model\new\tunisian_asr\tunisian.arpa
----5---10---15---20---25---30---35---40---45---50---55---60---65---70---75---80---85
**********************************************************************************
```

Pros and drawbacks of the customized ASR Solution using BrainSpeech and Tunisian Language Model:

T-ASR (Pros)	T-ASR (Cons)
• **Tailored to Tunisian Dialect:** A customized solution is designed specifically for the nuances of the Tunisian dialect, potentially leading to higher transcription accuracy (depending on the language model). • **Customization**: With the right expertise, the solution can be tailored to meet specific needs and preferences. • **Precision:** A very precise Tunisian speech to text conversion +90%.	• **Development Complexity:** Custom solutions can be more complex to develop and maintain compared to third-party APIs. • **Resource Requirements:** Developing and training a customized ASR solution demands significant computational resources and expertise. • **Current state:** In the current state of the ASR, the model does not recognize foreign languages and dialects other than Tunisian dialect. • ***Execution requirements:*** Requires a custom virtual environment for execution.