July 2023

ACCENT ADJUSTMENT BASED ON SPOKEN FEEDBACK

Keith Griffin

ACCENT ADJUSTMENT BASED ON SPOKEN FEEDBACK

AUTHOR:
Keith Griffin

ABSTRACT

In some instances, such as for online video conferences, call center interactions, etc., it can be difficult to understand what someone is saying based on their accent. Presented herein is a solution through which natural language processing (NLP) can be utilized to facilitate an accent detector that can determine a problematic accent intent and adjust a speaker's accent by adjusting and normalizing the speaker's speech to a more relatable accent of a person with whom the speaker is communicating.

DETAILED DESCRIPTION

Not being able to understand what someone is saying based on their accent is a common problem for people communicating across all walks of life and also in business, such as for in-person meetings, video conferences, calls, contact center interactions, and/or the like. The problem is often amplified when organizations move customer care and contact center services to locations resulting in callers interacting with agents that are non-native speakers of the callers' language. Such scenarios can result in frustration for a caller and callee and, in some instances, can impact an enterprise's business. For example, contact center businesses may incur high costs for agent accent training in some parts of the world. Thus, it would be advantageous to provide an automated solution for adjusting accents of one or more parties to a call.

Proposed herein is a potential solution to address such issues by providing an accent detector that can utilize NLP in order to react to a negative intent detected in the speech/accent of one party of a call (e.g., through spoken feedback, etc.) and adjust the accent of one of the parties to be a more relatable accent of the other party (e.g., the caller's accent or the callee's accent). The solution as proposed herein can be utilized across different scenarios, such as for contact centers, calling solutions, meeting solutions, and/or the like.

Figure 1, below, illustrates an example architecture for a system through which techniques proposed herein for providing accent adjustment based on spoken feedback may be implemented.

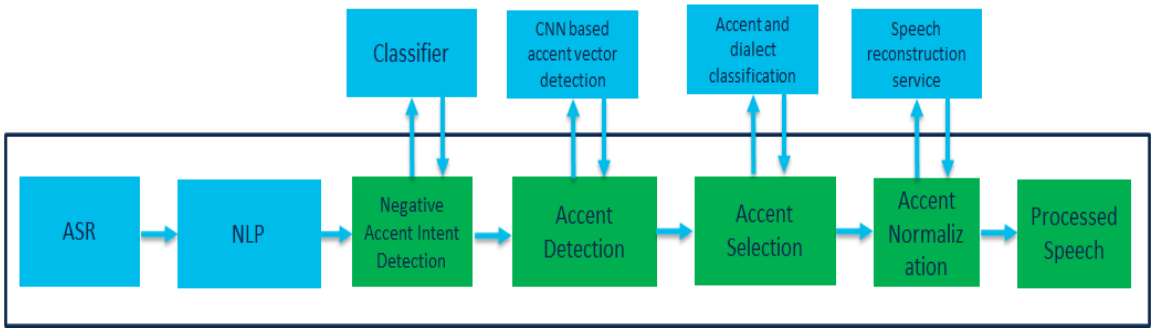

Figure 1: Example System Architecture to Facilitate Accent Adjustment Based on Spoken Feedback

As illustrated in Figure 1, the example system architecture may include various core and speech dataflow components, such as a speech and NLP system, an accent detection service, an accent selection service, and an accent normalization service, with new components shown in green. Some services may be implemented as cloud or client-based services. For example, the speech and NLP services may be cloud-based, while the accent detection and selection services can be cloud or client-based, and the accent normalization service can be cloud or client based with client-based service being the preferred architecture for real time performance reasons.

It should be noted, that for limited spoken sample sets a small client-side keyword Automatic Speech Recognition (ASR)/ Natural Language Understanding (NLU) could potentially be utilized, which would result in an entirely client-based architecture. Similarly, an entirely cloud-based architecture is possible. The optimal architecture for accuracy and performance is design with ASR/NLP and accent detection as cloud services, with accent adjustment on the client-side.

Consider an example operational flow that may be facilitated via the system of Figure 1. For example, consider that a caller and a callee are involved in an exchange in which the callee questions the caller's accent (e.g., "I'm very sorry but I can't understand what you are saying.").

In this example, the speech / NLP system can derive a negative accent intent (e.g., accent understanding), based on the callee's comment. Thereafter, the accent detection service is used to detect each of the caller's accent and the callee's accent, say, for example, detecting an English-Irish accent for the caller and an English-US accent for the callee.

The accent selection service is used to select the closest-matched normalization. For example, in the present scenario, the English-Irish accent could be normalized to an English-US accent such that the accent normalization service acts on the caller's accent to modify spoken communications from the caller to match the accent type of the callee (English-US).

It should be noted that accent normalization is known and thus, is not considered a novel aspect of the solution proposed herein. In at least one implementation, speech detection may be provided using a Convolutional Neural Network (CNN) that is on vectorized speech using a technique, such as Mel Frequency Cepstral Coefficients (MFCC), in order to determine the characteristics of speech for a given accent.

In some instances, accent adjustment can use a related approach by adjusting the speech output to the characteristics of a given accent. In some instances, the speech / NLP step can be skipped with direct control provided to the caller to invoke the service where accent detection and matching would occur as a result.

Accordingly, a solution is provided herein through which an accent detector can be provided via NLP in order to determine a problematic accent intent and adjust a speaker's accent by adjusting and normalizing the speaker's speech to a more relatable accent of a person with whom the speaker is communicating. Although the example system flow and accent normalization as discussed above is relatively mild, it is to be understood that techniques as presented herein may be capable of normalizing larger variations in accent.