# A personalised speech communication application for dysarthric speakers

*Matthew Gibson*[1], *Ievgen Karaulov*, *Oleksii Zhelo*[1], *Filip Jurcicek*[1]

[1]Voiceitt

matt,oleksii,filip@voiceitt.com

## Abstract

Individuals with impaired speech are often understood only by those familiar with their speech e.g. a care-giver or close family member. These impaired speakers are therefore highly dependent upon those familiar listeners for their spoken communication needs. These needs vary from basic expressions of hunger or thirst to much more advanced requirements like being understood at a work meeting. A significant subset of individuals with impaired speech also have reduced motor function which limits their mobility or dexterity. For this subset of individuals, the ability to communicate via the medium of speech is crucial. This paper describes a personalised speech communication application targeted towards English language speakers with impaired speech. This application enables the user to hold conversations with other humans, dictate text to a machine and participate in meetings via closed captioning.

**Index Terms**: speech recognition, dysarthric speech

## 1. Introduction

Atypical speech related to an underlying medical disease or disability is often clinically referred to as dysarthria [1]. Dysarthria is associated with developmental disorders (e.g. cerebral palsy, Down's syndrome), degenerative illness (e.g. Parkinsons, multiple sclerosis), traumatic brain injury and stroke. A subset of individuals with dysarthria also have reduced mobility or dexterity related to their underlying condition, and would benefit more than the average user from accurate human-machine voice interfaces.

However, despite significant advances in the field of automatic speech recognition (ASR) over the past fifteen years, off-the-shelf speech technology remains too inaccurate to be useful for many dysarthric speakers. Dysarthric ASR is therefore an area of ongoing research ([2, 3, 4, 5, 6]). Key to this research is the idea of personalization, adapting the technology to better match the characteristics of the user. Acoustic model adaptation is a particularly powerful form of personalization which has improved ASR accuracy for many dysarthric speakers. The application described in this paper leverages speaker adaptation of deep neural network (DNN) based acoustic models to enable dysarthric speakers to communicate via multiple channels.

## 2. System description

An illustration of the system is given in Figure 1. The front-end is a web application which runs within a browser on an edge device; a desktop machine, laptop, tablet or mobile phone. The back end is a large vocabulary, personalized speech recognition system based upon finetuning of state-of-the-art deep neural network acoustic models to individual speakers.

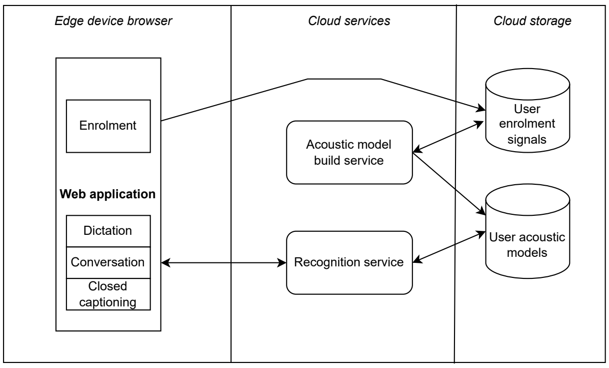


Figure 1: *System overview.*

This application initially guides the user through a registration and enrolment phase, where they provide samples of their speech by recording a set of prompts. Figure 2 is a screenshot of the enrolment phase using the mobile application. These speech signals and associated transcriptions are stored for subsequent acoustic model adaptation.

Once a sufficient quantity of enrolment signals have been received, the back end acoustic model build service generates a user-specific acoustic model. This is a relatively compute-intense process which requires cloud compute and, more specifically, access to graphical processing units.

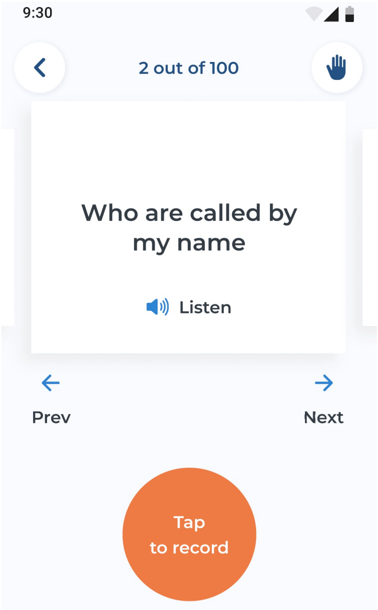
Once a user-specific acoustic model is generated the dictation, conversation and closed captioning modes are enabled on the web application. The text-to-speech functionality required by these modes is provided by the cloud-based recognition service, reducing the compute and memory requirements of the edge device.

## 3. Application functionality

The application is designed to enable users to communicate in several ways and has multiple modes to reflect this; dictation, conversation and closed captioning.

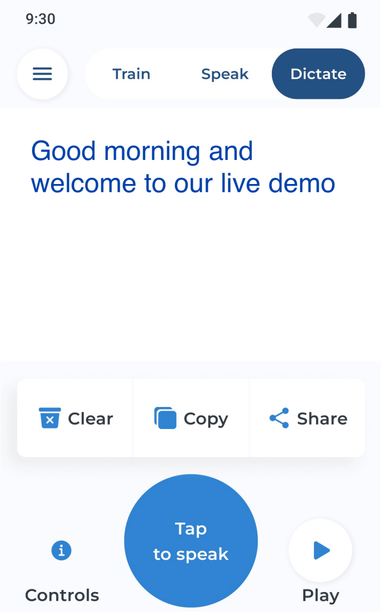
### 3.1. Dictation

Dictation mode is designed for users who wish to use speech to generate electronic text, enabling the user to dictate documents, emails or text messages. This is particularly useful for users with limited mobility or dexterity. The application converts the user's speech to text, which is subsequently available for editing, copying and sharing to an external application e.g. a messaging or email application. Figure 3 is a screenshot of the application in dictation mode on a mobile phone device.



Who are called by my name

Figure 2: *Enrolment with the mobile application.*



Good morning and welcome to our live demo

Figure 3: *Mobile application in dictation mode.*

In dictation mode, support for spoken punctuation commands ( e.g. 'question mark', 'comma' ) and text formatting commands ( e.g. 'new line' ) is provided.

### 3.2. Conversation

Conversation mode enables the user to conduct in-person speech communication with another human. The user's speech is converted to text and the recognised text is subsequently re-synthesised in a synthetic voice. The re-synthesised speech is typically more intelligible than the user's speech, particularly for listeners who are unfamiliar with the user's speaking style. The user can customise the synthetic voice to their preferred gender, accent and style.

An additional feature of conversation mode is the shortcut phrase. A shortcut phrase is a shortened version of a longer phrase e.g. 'Coffee' is a shortened version of 'May I have a coffee please?'. In conversation mode, the user utters the shortened version of the phrase and the application synthesises the longer version. This reduces the time and effort required by the user, an important consideration for those users for whom speech is more effortful.

### 3.3. Closed captioning

Closed caption mode enables the user to participate in electronic meetings e.g. via Microsoft Teams or Zoom. Automatic closed caption transcriptions of the user's speech are generated as the user speaks, allowing the user to communicate with meeting participants with the convenience of speech.

## 4. Conclusions

The communication application described above enables users with impaired speech to communicate more effectively via multiple channels. The idea behind the show-and-tell submission is to provide a live demonstration of the application's performance and functionality with a dysarthric user.

## 5. References

[1] J. R. Duffy, *Motor Speech Disorders. Substrates, Differential Diagnosis, and Management*. Philadelphia: Elsevier Mosby, 2005.

[2] J. Shor, D. Emanuel, O. Lang, O. Tuval, M. Brenner, J. Cattiau, F. Vieira, M. McNally, T. Charbonneau, M. Nollstadt, A. Hassidim, and Y. Matias, "Personalizing ASR for dysarthric and accented speech with limited data," in *Proc. Interspeech*, 2019.

[3] J. Green, R. MacDonald, P.-P. Jiang, J. Cattiau, R. Heywood, R. Cave, K. Seaver, M. Ladewig, J. Tobin, M. Brenner, P. Nelson, and K. Tomanek, "Automatic speech recognition of disordered speech: Personalized models outperforming human listeners on short phrases," in *Proc. Interspeech*, 2021.

[4] K. Tomanek, V. Zayats, D. Padfield, K. Vaillancourt, and F. Biadsy, "Residual adapters for parameter-efficient ASR adaptation to atypical and accented speech," in *Proc. EMNLP*, 2021.

[5] M. K. Baskar, T. Herzig, D. Nguyen, M. Diez, T. Polzehland, L. Burget, and J. Černocký, "Speaker adaptation for Wav2vec2 based dysarthric ASR," in *Proc. Interspeech*, 2022.

[6] L. P. Violeta, W.-C. Huang, and T. Toda, "Investigating self-supervised pretraining frameworks for pathological speech recognition," in *Proc. Interspeech*, 2022.