Using Artificial Intelligence to assist with the understanding of speech of people living with Parkinson's Disease.

Paul Gadd

Protocol IT Ltd, Advanced Technology & Innovation Centre, Loughborough, UK

Protocol IT ⇄
We speak your language

Introduction

Parkinson's disease (PD) is a condition in which parts of the brain become progressively damaged over many years and is projected to continue to rise in incidence, being driven by more factors than just ageing. Issues where neurological damage affects the quality of speech, can result in further withdrawal as communication becomes more difficult, leading to social isolation.

The aim of the research is to develop a tool to assist people living with PD where speech has been affected and enabling them to be understood, using Artificial Intelligence to convert the speech into text that can be displayed, to assist with understanding.

The research has been focused on capturing voice recordings from participants diagnosed with PD, which are used to fine-tune or re-train an existing artificial intelligence language model. This is subsequently used to translate speech to text with additional interpretation from Natural Language Processing (NLP).

Research description

In June 2022 a call for participants was launched through Parkinson's UK for face-to-face recordings. This was expanded in January 2023 to include online recordings via Zoom. Since sampled speech is required to be optimised at 16kHz for re-training existing language models, Zoom recordings would have minimal effect on Automatic Speech Recognition (ASR), but would significantly increase the number of participants able to contribute.
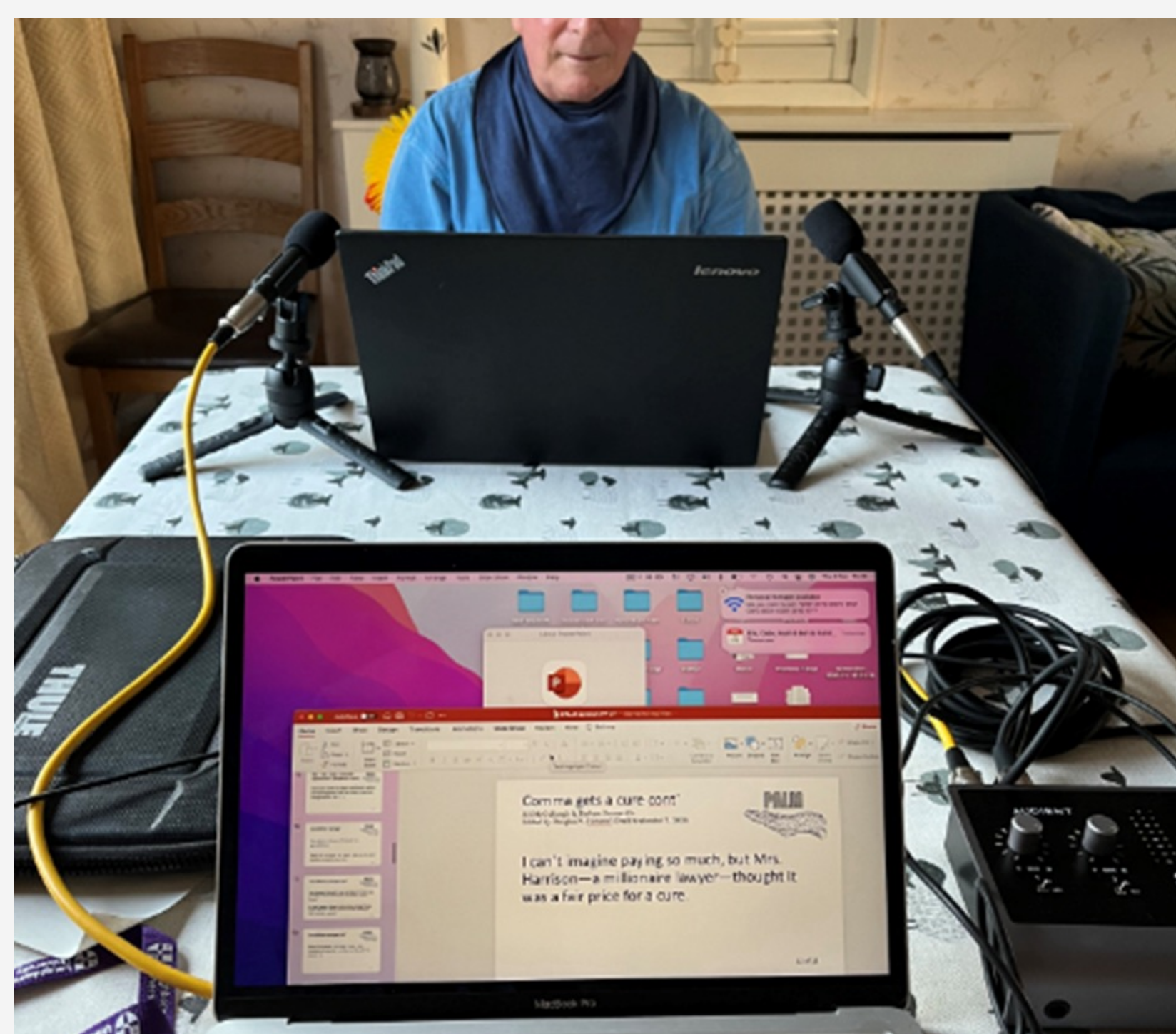
The research methodology involves gathering sampled speech using matched pair microphones. Four recording stages are used for participants:
1) Standard text passages[1], often used by speech therapists for linguistic analysis, are read. These passages assist with ease of labelling the speech files.
2) 'Free speech' on a subject chosen by each participant.
3) Selection of short sentences[2]
4) Vocal exercises[3] interspersed with short read passages.

Fig. 1: Photo from a face to face recording session

Previous research at Kings College in London[4] and at Università degli Studi di Bari 'Aldo Moro' [5] helped focus the selection of reading material used during speech recordings.

Where possible, recordings are made at a specific time of the day when participants find their speech most affected.

The Nvidia ASR toolkit was selected as the platform for research, due to the low Word Error Rate (WER), providing a higher percentage of accuracy. The recorded speech is used with labelling to fine-tune or re-train an existing acoustic language model.

Results

Initial assessment to determine whether existing language models are capable of accurately performing ASR, are carried out by feeding recorded samples passing through a local instance of the Nvidia trained model. This incorporates NLP which then tries to make sense of the recording into natural language.

Fig 2. shows the stereo samples of a participant's speech, significantly impacted as a result of neurological damage, reading one of the passages, resulted in a high WER.
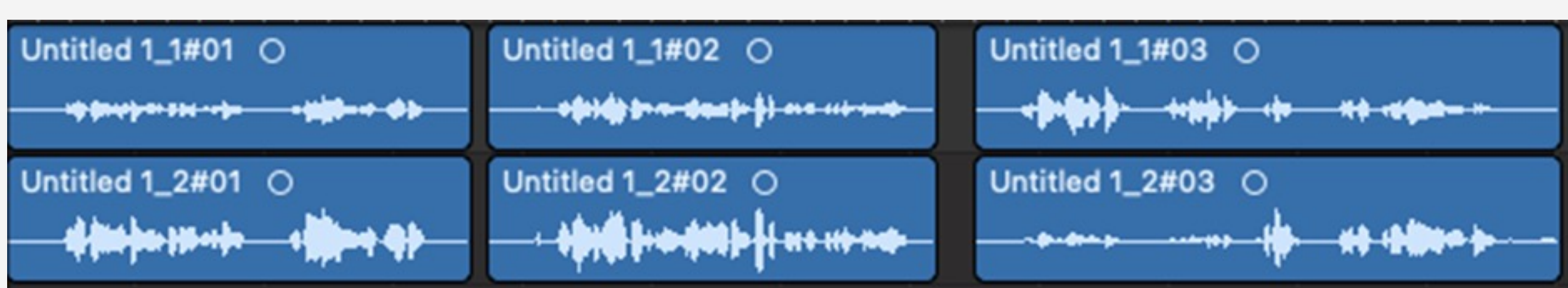

Fig. 2: Recorded pre-determined speech in short 12-15 second slots in stereo.

Results continued

When the speech is passed through an existing Nvidia language model, it results in a high WER. The first section shows the speech manually transcribed, followed by Nvidia ASR transcribed.

Manual transcribed speech:
One night the rats heard a loud, a loud noise in the loft. It was a very dreary old place. The roof let the rain come washing it in, washing the beams and rafters had all rotted through, so that the whole thing was quite unsafe.

Nvidia ASR transcribed speech:
On that. Loud noise in the loft. It the only ju place the roof about the rain, com in, walking in, walking the reams and ras at all right through. The whole thing was quite unsafe.

Re-training a neuro-network language model with additional audio sample with associated labels, would significantly decrease the WER, and therefore accuracy of the ASR.
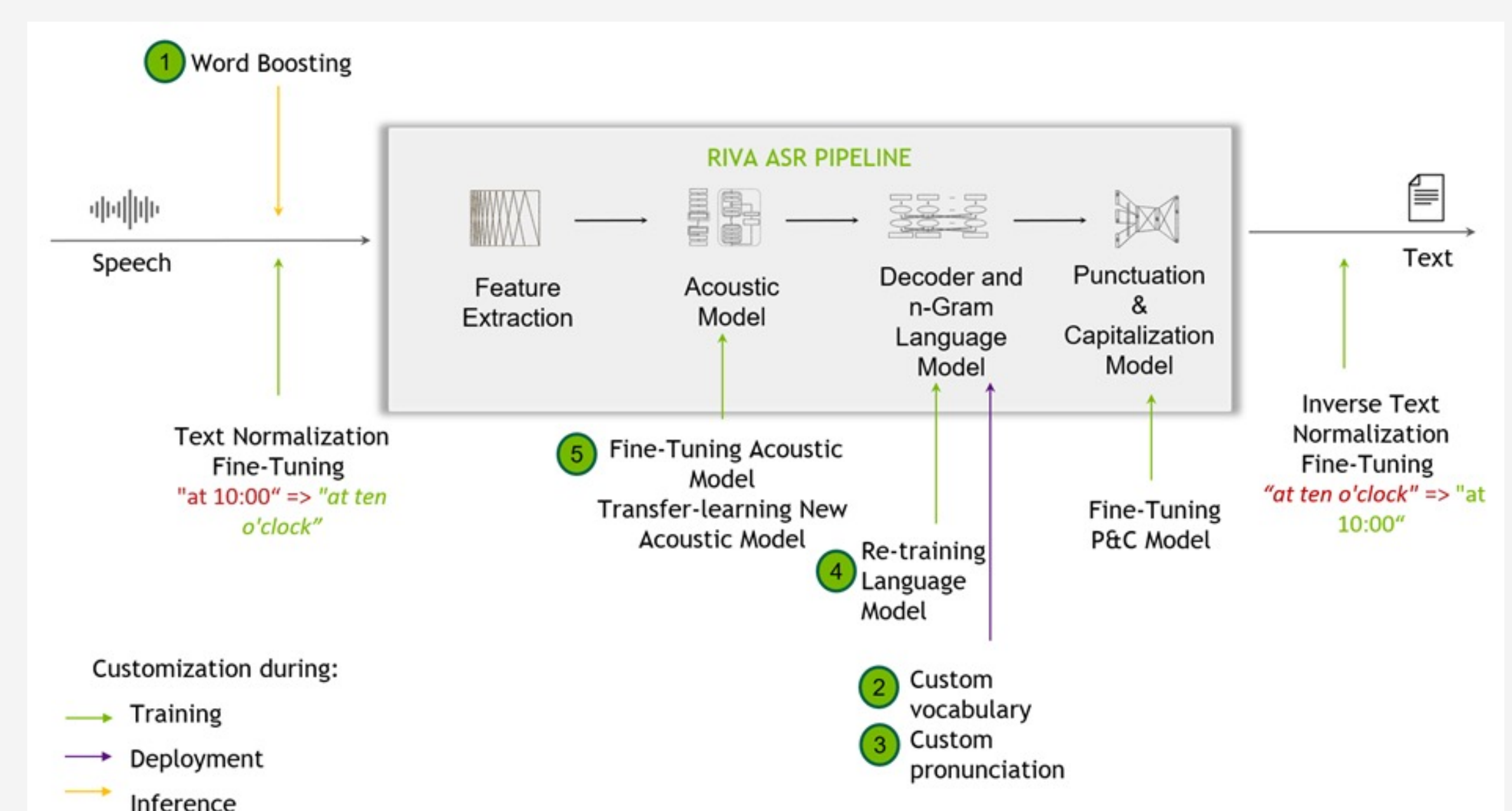

Fig. 3: Detailing the Nvidia re-training speech model

For participants where speech still converts with a low WER, it was still considered worthwhile to record using the same techniques. The data then provides a reference point to compare with future recordings to assist with predicting deterioration. This would be used to re-train the language model without further speech samples.

Conclusions & impact

The project is still in the data collection phase. Early results illustrate that neurological changes resulting from PD affect the intelligibility of speech in people in different ways, and existing language models are not able to accurately translate speech to text.

We are currently looking to expand links with speech therapists and other support groups to obtain additional speech recordings, for people with significantly impacted speech by PD and other neurological conditions. This will help expand the number of eligible participants and therefore help re-train a more diverse language model.

An unexpected outcome from recording the vocal exercises, was motivating participants to engage with regular exercise, after discussing the visual and auditory speech samples. This is another area we are keen to explore with a goal of developing other AI tools to form part of the speech therapy toolkit.

How can you help

Volunteers
We are looking for other sources who can signpost potential volunteers with PD where speech is significantly affected. If you can support us increase the speech samples, please get in touch.

Re-training models by predicting speech deterioration
We are looking to work with other researchers to investigate predicting changes in re-trained language models as a result of speech deterioration. Please get in touch if this is of interest.

References

1. Standard linguistic resources provided by York University for students:
https://www.york.ac.uk/media/languageandlinguistics/documents/currentstudents/linguisticsresources/Standardised-reading.pdf
2. Short sentences taken from the Google Euphonia project:
https://sites.research.google/euphonia/about/
3. Speech exercise sheet developed by the Parkinson's Foundation
https://www.protocolit.co.uk/wp-content/uploads/2023/06/Exercising-Your-Speech-and-Voice-System.pdf
4. Mobile Device Voice Recordings at King's College London (MDVR-KCL) from both early and advanced Parkinson's disease patients and healthy controls - May 17, 2019
H. Jaeger; D. Trivedi; M. Stadtschnitzer [DOI : 10.5281/zenodo.2867216]
5. Assessment of Speech Intelligibility in PD Using an STT System – ISSN : 2169-3536 IEEE 17th Oct 2017;
G. Dimauro, V. di Nicola, V. Bevilacqua, D. Caivano, and F. Girardi. [DOI ;10.1109 / ACCESS.2017.2762475]

info@protocolit.co.uk
https://www.protocolit.co.uk/ukspeech-2023

UK Speech

UK SPEECH 2023 SHEFFIELD

Protocol IT ⇄
We speak your language