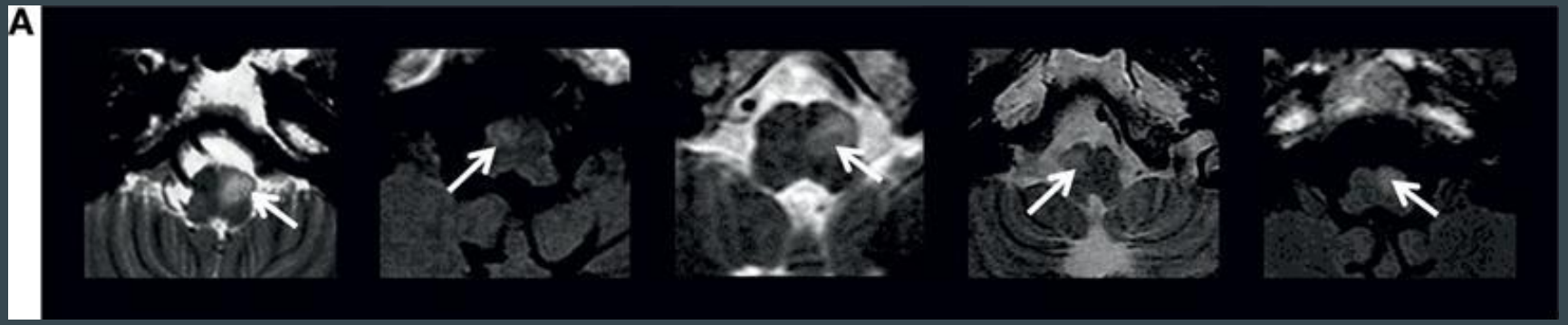
Speech-to-Text and Speech-to-Speech for People Affected by Scanning Speech

...

Gracie Richards, Gabriel Rodriguez, Pratiksha Pradhan

Hypertrophic Olivary Degeneration

- A disease that has a rare chance of occurring after a brain hemorrhage, tumor, injury, stroke, cavernous malformation, or brain surgery
- Can also be idiopathic, with no known cause
- Characterized by degeneration of the olives in the brain
  - The olives are essential for the cerebellum, the part of the brain responsible for motor control, to function
- As a result, HOD affects nearly all aspects of motor function, or movement and coordination



White arrows indicate abnormal inferior olives.

Tilikete C, Desestret V. Hypertrophic Olivary Degeneration and Palatal or Oculopalatal Tremor. Front Neurol. 2017 Jun 29;8:302. doi: 10.3389/fneur.2017.00302. PMID: 28706504; PMCID: PMC5490180.

Scanning speech and its challenges

- HOD can result in scanning speech, “a type of ataxic dysarthria in which spoken words are broken up into separate syllables, often separated by a noticeable pause, and spoken with varying force”[1]
- Scanning speech can be found in many conditions, including:
  - Multiple sclerosis
  - Stroke
  - Brain tumors
  - Brain injury
  - Parkinson’s disease
  - ALS
  - Huntington’s disease
  - Cerebral palsy
  - Muscular dystrophy
  - HOD

Stroke survivor speaks about difficulties with his speech:

[1]Scanning speech [Internet]. Wikimedia Foundation; 2024 [cited 2024 Mar 3]. Available from: https://en.wikipedia.org/wiki/Scanning_speech

App functionality

- Speech-to-text - for in-person interactions
  - The phone is shown to the person the user is talking to
  - Captions in real-time what the user is saying
- Speech-to-speech - for messaging
  - The user creates audio recordings of their voice and the app enhances it
    - Removes extraneous pauses and speeds up the speech
  - Can then save or send the audio recording
  - There is the option to have the recording come with a text transcription
  - Has the advantage over texting in that it allows them to communicate with their own voice
- Preset messages
  - Can save messages for frequent use
  - E.g. “Hi my name is ___, I have a disorder called HOD which affects my movements and coordination.”, or their address or phone number.

Discussion

- The motor difficulties caused by HOD make speaking a challenge, and it becomes harder for others to understand HOD patients
- Although HOD seems to disrupt functionality in all parts of the body, cognition is unaffected
  - They can think perfectly well, communicating their thoughts is the problem
- Our app aims to help people with HOD and other conditions who have scanning speech communicate more easily
- The design is speech-to-text and speech-to-speech components trained to recognize scanning speech
- Can be used for in-person and virtual interactions
- Preset messages allow the user to communicate their most commonly needed phrases with little effort

Challenges

- Finding a solution that will help people with HOD
  - Most aspects of HOD are a mystery – cause is unknown, no treatments
  - No known genetic cause, and not enough data available
  - Limited data on HOD available online
    - Partly because it's such a rare disease
    - Few videos of HOD speech available
    - Only individual case studies are available, no such centralized database
  - Cannot answer any biological questions in this hackathon (e.g. what cell types are involved in causing the disease, what genes increase risk), so we decided to focus on something that affects patients' daily life
- Implementation
- Lack of openly available data out there on scanned speech- cannot use data samples to implement our methodology and create a model based on it

Future Directions

- Firstly, creating a HOD-specific database that includes patient clinical and imaging symptoms, treatments, past medical history, scanned speech samples, etc. and encouraging information sharing among organizations
- Create the algorithm that trains on scanning speech and is able to recognize it
- Build the web app, taking steps to make it accessible to people with motor difficulty
- Incorporate more advanced speech enhancement beyond removing extraneous pauses and increasing speech speed, using signal processing and/or machine learning
- Can implement multimodal speech-to-text recognition, incorporating lip movements, to improve accuracy of scanned speech recognition
- Allow real patients to use the app and improve it based on their feedback