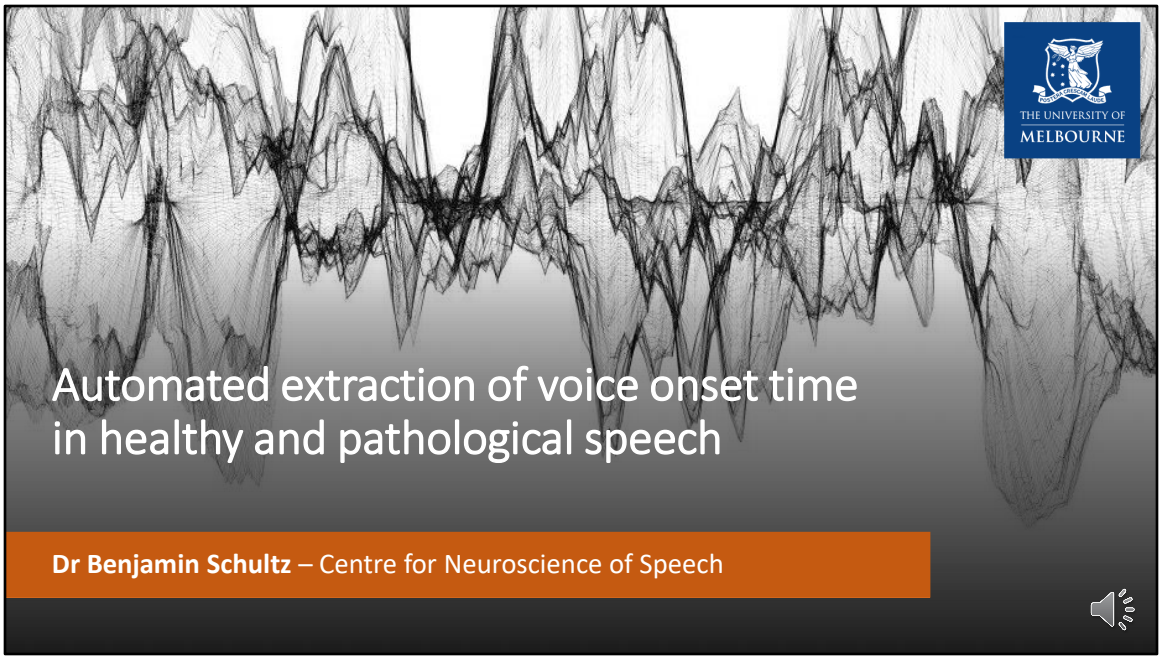
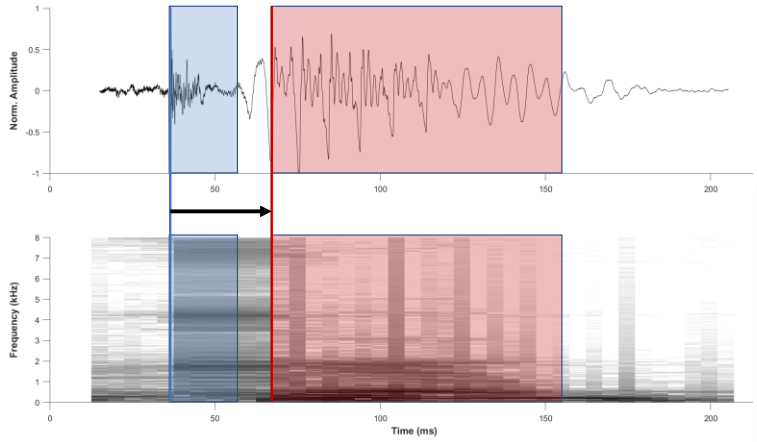

THE UNIVERSITY OF
MELBOURNE
Automated extraction of voice onset time in healthy and pathological speech
Dr Benjamin Schultz – Centre for Neuroscience of Speech

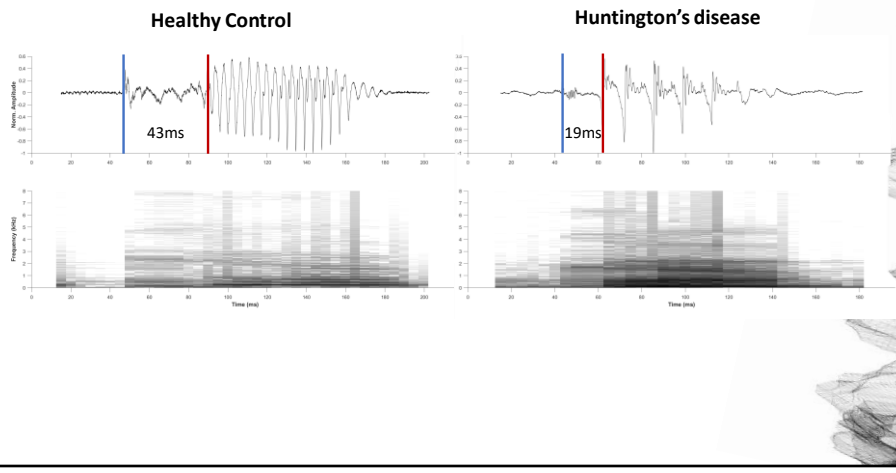
Voice Onset Time

Time between **burst** of a consonant and **voicing** of vowel



Voice onset time is a measure of articulatory timing that measures the latency between the burst of a consonant (shown here in blue) and the voicing of the following vowel (shown here in red).

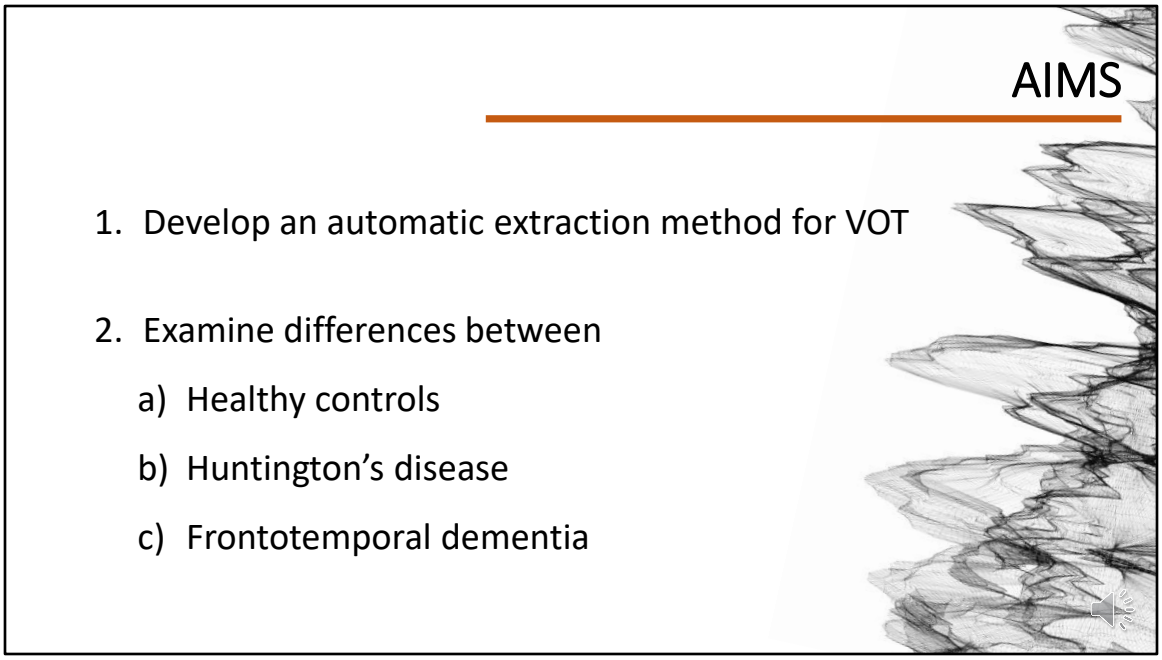
VOT & Neurodegenerative Disease



Some studies have shown that VOT differs between healthy controls and groups with neurodegenerative disease.

For example, healthy controls have shown longer VOTs than people with Huntington's disease. Similar effects have been found for people with Parkinson's disease.

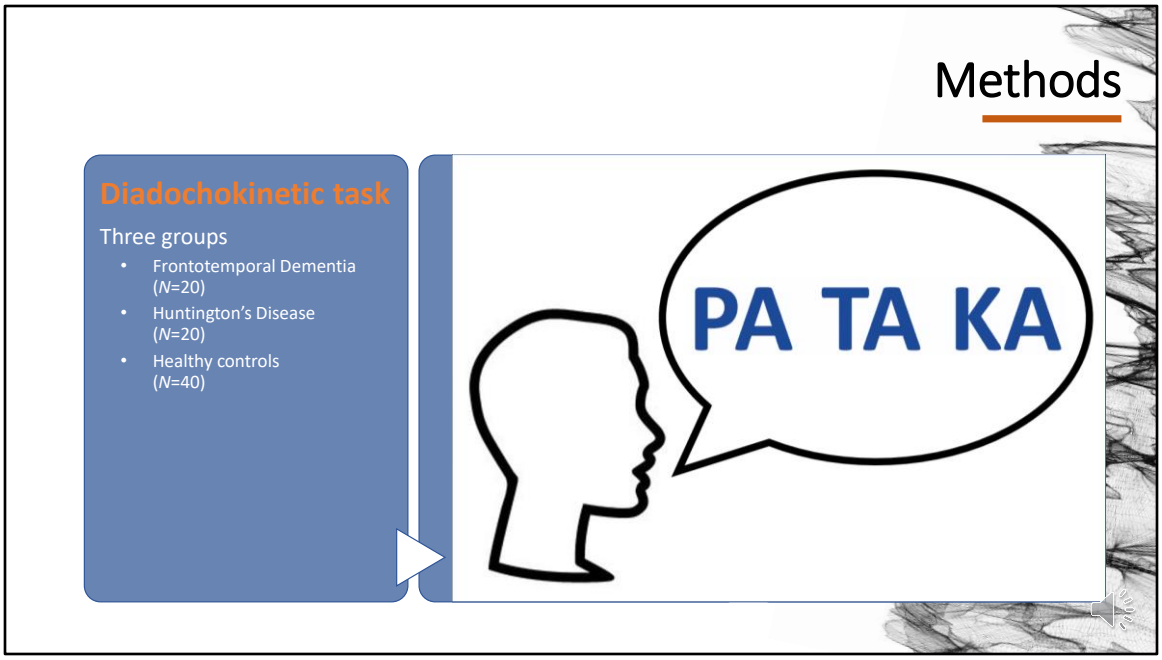
Therefore, VOT could be a valuable biomarker for the detection of neurodegenerative disease and other conditions that affect speech.



Manual extraction of VOT can take a great deal of time.
It can be difficult to determine VOT for connected speech

To resolve these issues, we developed an algorithm for the automatic extraction of VOT for continuous, unsegmented speech data.

We then assessed whether this algorithm could be used to distinguish neurodegenerative disease from healthy speech.



Speakers were people with frontotemporal dementia, Huntington's disease, and healthy controls.

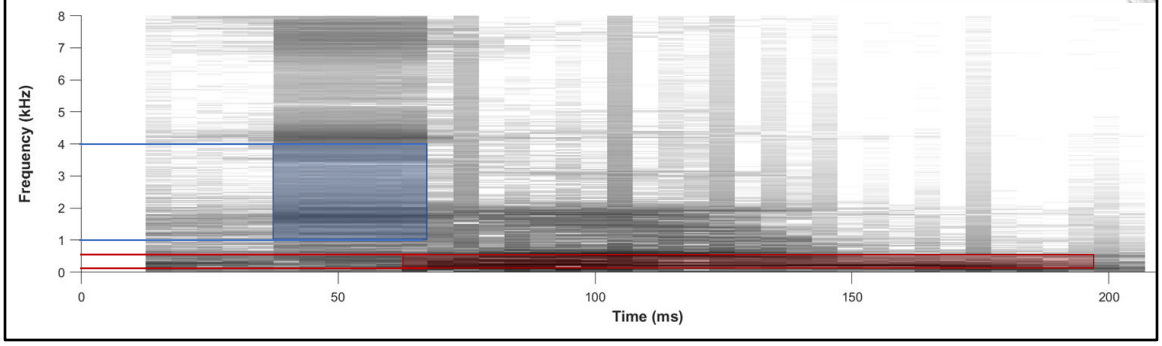
Speakers performed a diadochokinetic task where they repeated the syllables /pa/, /ta/, and /ka/

Unsegmented audio recordings were subjected to the VOT extraction algorithm

VOT was also manually annotated by two independent raters using Praat software

VOT Extraction Algorithm

1. Detect **bursts** using high-frequency summed energy (1000Hz-4000Hz)
2. Detect **voicing** using low-frequency summed energy (75Hz-500Hz)

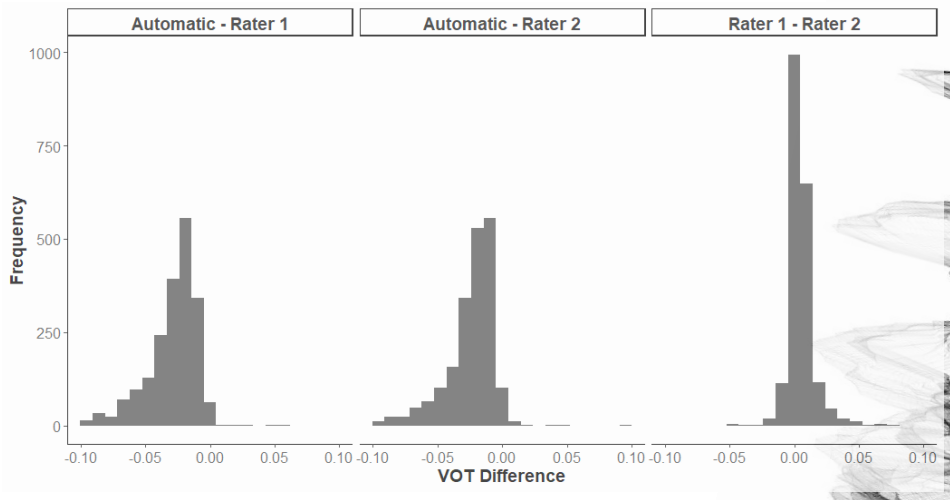


First, syllable onsets and offsets were determined using summed energy from 75Hz to 8000Hz.

The VOT algorithm determined bursts using the summed energy across frequencies between 1000Hz and 4000Hz, as shown in the blue section here.

Voicing was determined using the summed energy across frequencies considered to reflect the fundamental frequency, specifically 75Hz to 500Hz, as shown in the red section.

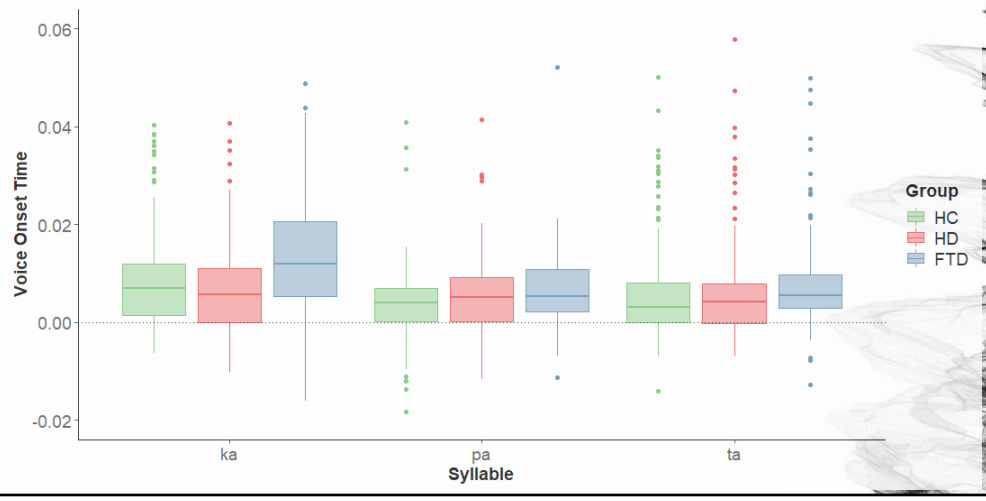
Results: Human raters vs. automatic extraction



Automatic VOT extraction underestimated VOTs relative to raters.

Burst times are more aligned than voicing onsets. Automatic VOT extraction detects voice onsets as occurring earlier than human raters.

Results: Group differences

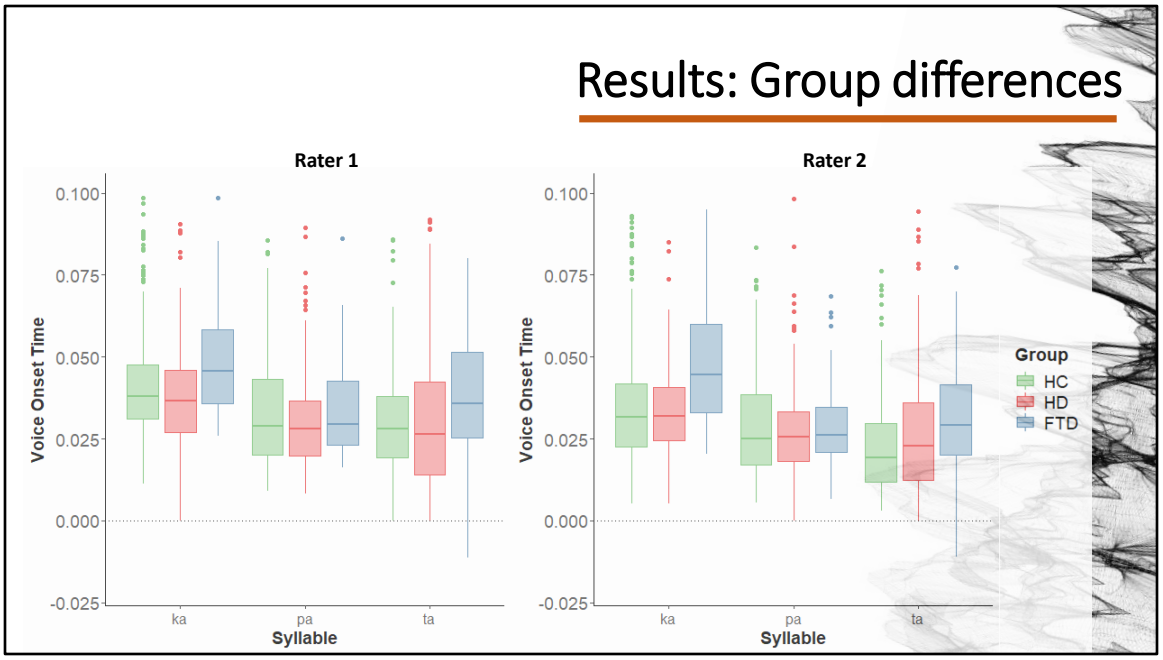


Linear mixed-effects model (VOT for automatic extraction)
Group F (2, 45.59) = 7.313, p = 0.00176 **
Syllable F (2, 101.88) = 5.203, p = 0.00706 **
Group:Syllable F (4, 94.75) = 1.324 p = 0.26680

Pairwise comparisons
HD – HC: p = 0.8382
FTD – HC: p = 0.0050 **
FTD – HD: p = 0.0011 **

pa – ka: p = 0.025 *
ta – ka: p = 0.700
ta – pa: p = 0.212

Results: Group differences

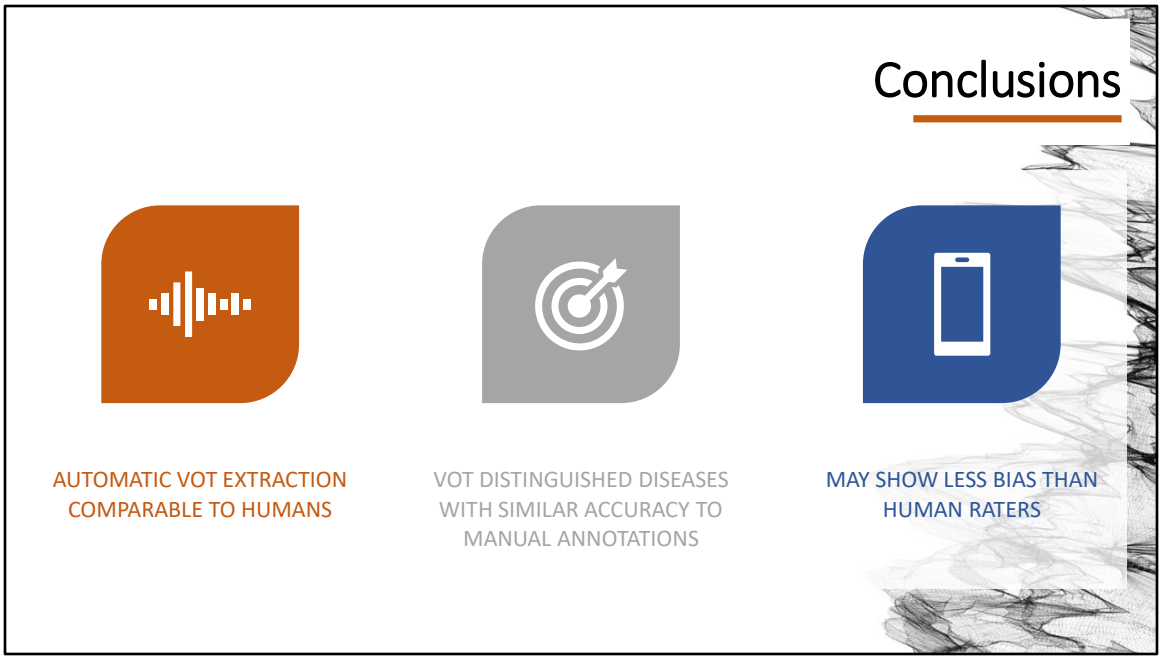


Linear mixed-effects model (VOT for Rater 1)
Group: F (2, 45.59) = 7.313, p = 0.00176 **
Syllable: F (2, 101.88) = 5.203, p = 0.00706 **
Group:Syllable: F (4, 94.75) = 1.324, p = 0.26680

HD – HC: p = 0.6638
FTD – HC: p = 0.0493 *
FTD – HD: p = 0.0065 **

Linear mixed-effects model (VOT for Rater 2)
Group: F (2, 43.32) = 1.773, p = 0.1819
Syllable: F (2, 93.21) = 21.887, p = 0.0000000161 ***
Group:Syllable: F (4, 91.77) = 3.067, p = 0.0202 *

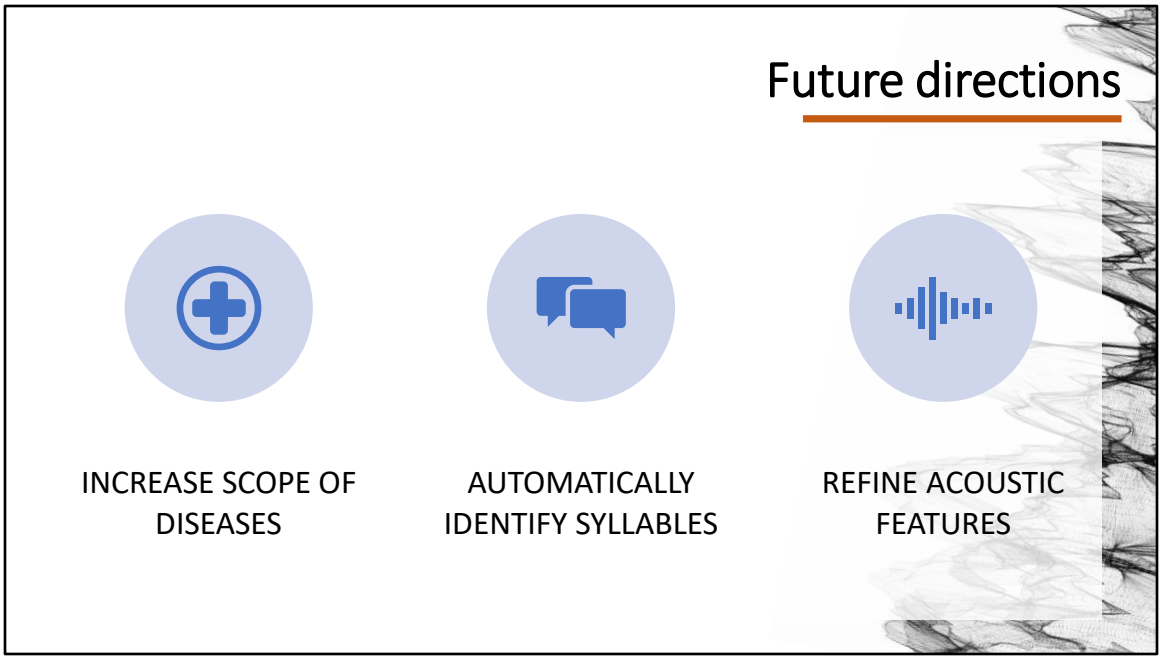
HD – HC: p = 0.8443
FTD – HC: p = 0.0210 *
FTD – HD: p = 0.0055 **



Our VOT extraction algorithm was comparable to that of humans.
Unclear whether human raters annotate voicing as too late or the algorithm finds voicing too early.

Regardless, automatic VOT extraction was able to distinguish FTD from HC and HD just as well as human raters.

The automatic VOT algorithm may show less bias than human raters but this needs to be examined more closely.



Examine Parkinson's disease, Friedreich's ataxia, and multiple sclerosis.

Include automatic detection of syllable labels (i.e., /pa/, /ta/, and/ka/) using supervised machine learning.

Determine acoustic measures that best capture voicing (e.g., f0 measures)

Acknowledgements

Centre for Neuroscience of Speech

- Adam Vogel
- Michelle Magee
- Gustavo Noffs
- Jess Chan
- Courtney Lewis

REDENLAB

Contact

Benjamin Schultz

Email: ben.schultz@unimelb.edu.au

Webpage: https://findanexpert.unimelb.edu.au/profile/856388-ben-schultz

THE UNIVERSITY OF MELBOURNE