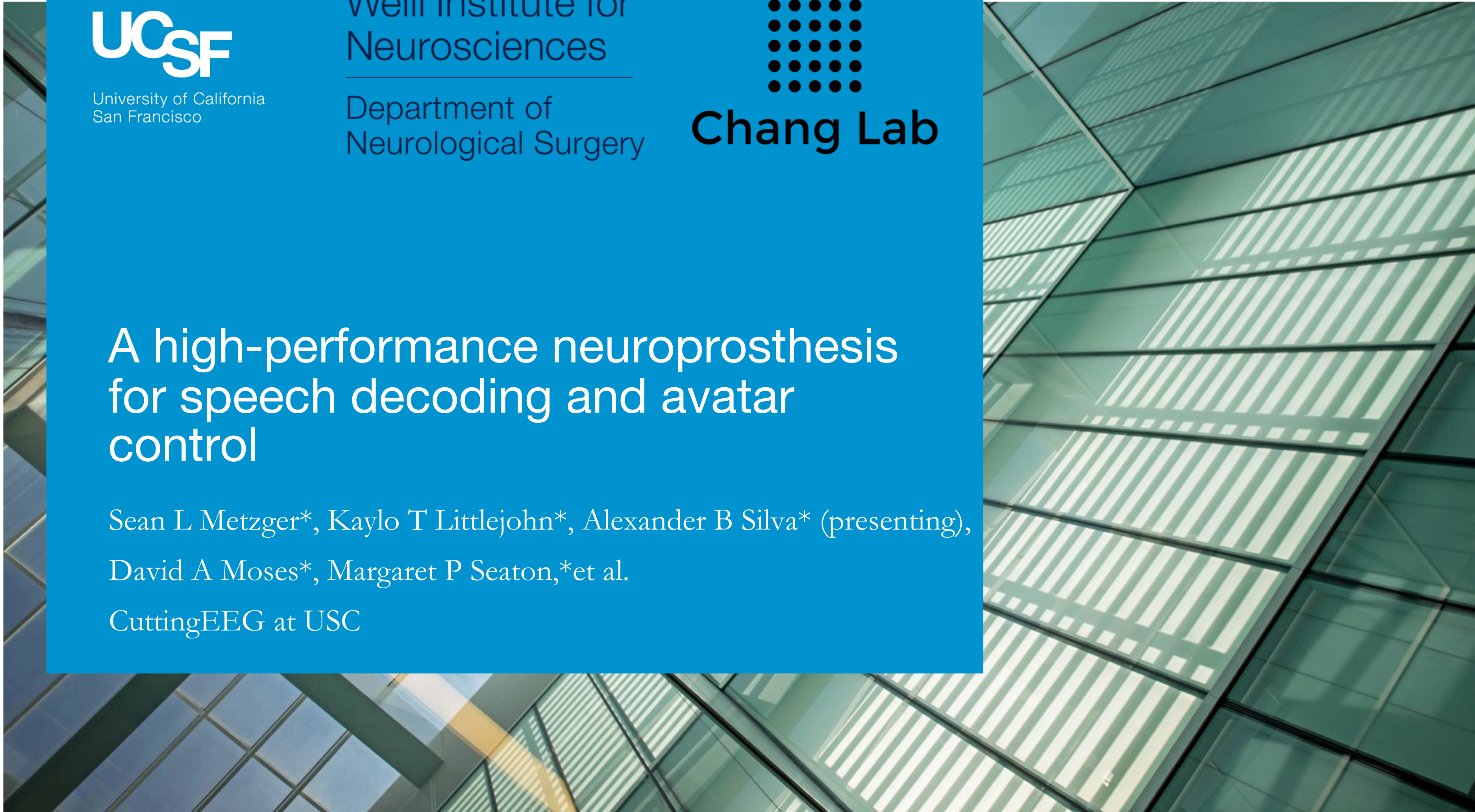


UCSF
University of California
San Francisco

Weill Institute for
Neurosciences

Department of
Neurological Surgery

Chang Lab

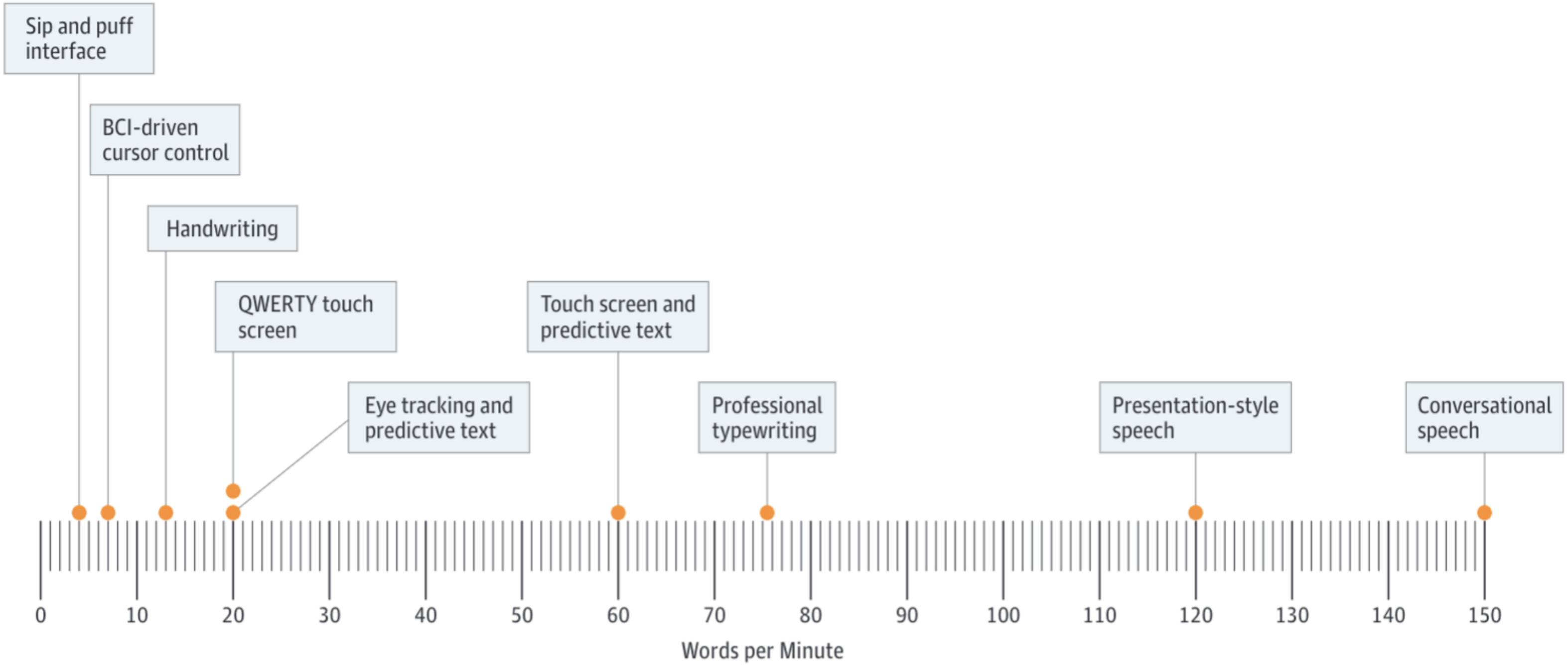
A high-performance neuroprosthesis for speech decoding and avatar control

Sean L Metzger*, Kaylo T Littlejohn*, Alexander B Silva* (presenting),

David A Moses*, Margaret P Seaton,*et al.

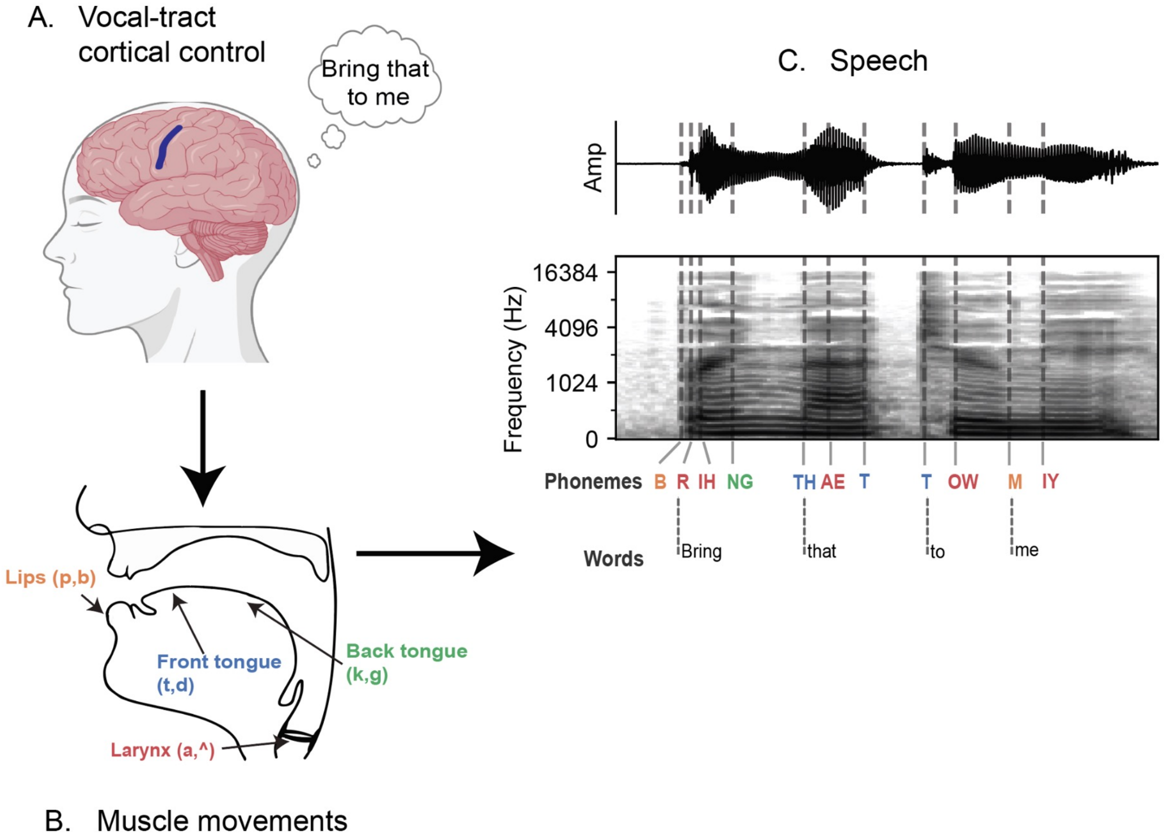
CuttingEEG at USC

Speech is a special form of communication



Chang EF, Anumanchipalli GK. Toward a Speech Neuroprosthesis. JAMA 2020

Overview of speech production

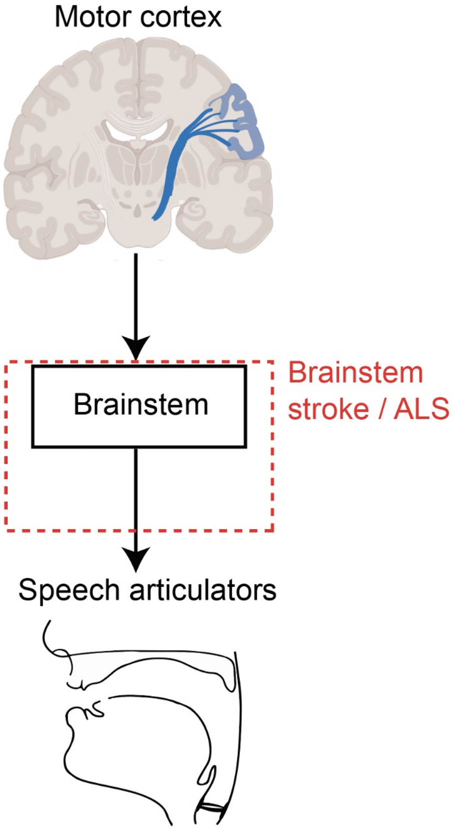


Speech articulators: muscle groups responsible for shaping the vocal-tract (jaws, lips, tongue, larynx)

Phonemes: smallest perceptually distinct sounds that form a language

Speech is multimodal: Words, sounds, and facial movements/expressions

Stroke and ALS cause loss of speech

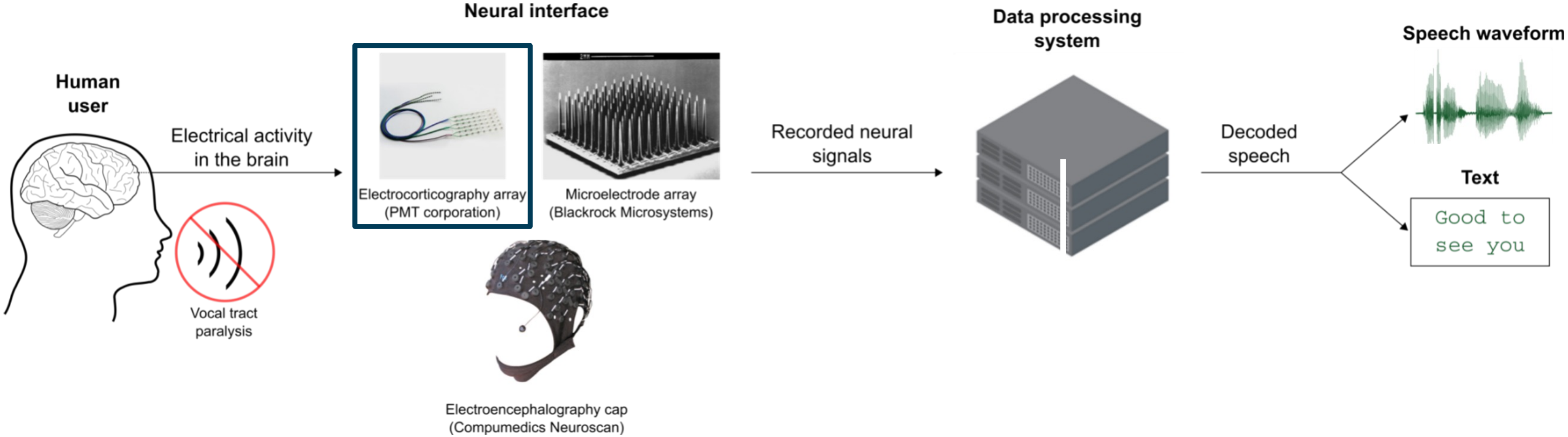


Anarthria:

- Inability to articulate speech
- Often co-occurs with loss of limb function

- ALS: ~5000 new cases a year (CDC)
- Brainstem stroke: 10-15% of all strokes (NIH)

A BCI could bypass diseased motor pathways to restore speech



Human user

Electrical activity in the brain

Vocal tract paralysis

Neural interface

Electrocorticography array (PMT corporation)

Microelectrode array (Blackrock Microsystems)

Electroencephalography cap (Compumedics Neuroscan)

Recorded neural signals

Data processing system

Decoded speech

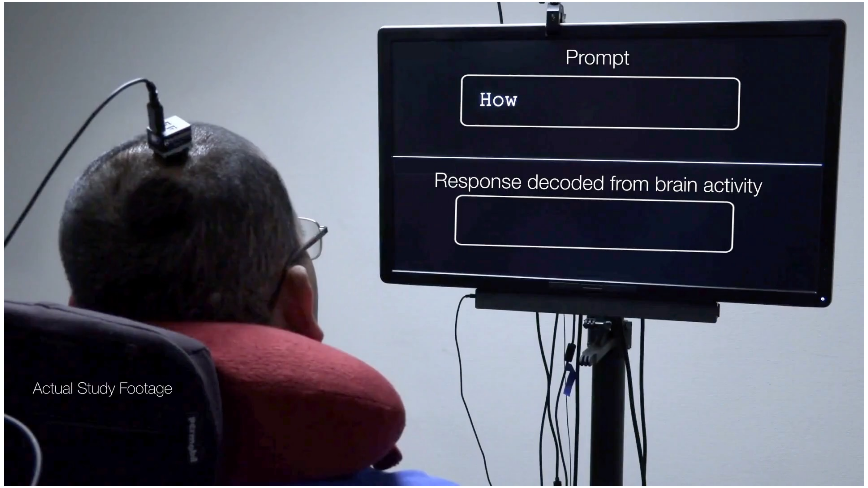
Speech waveform

Text

Good to see you

Embodied communication is multimodal

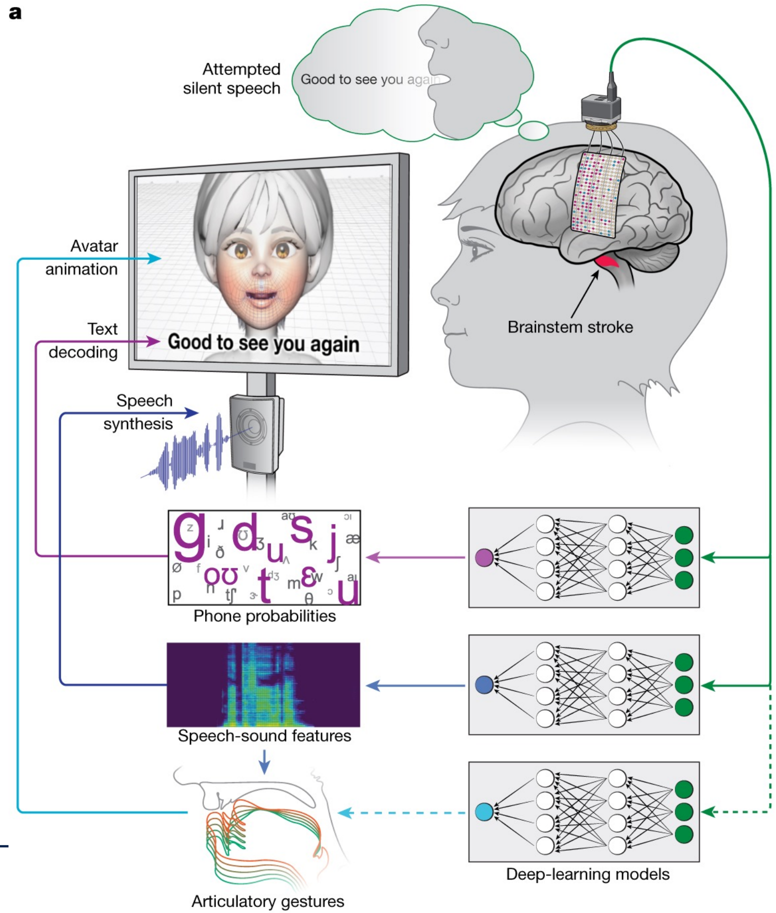
- Words and phrases
- Sounds : pitch and intonation
- Orofacial movements and expressions



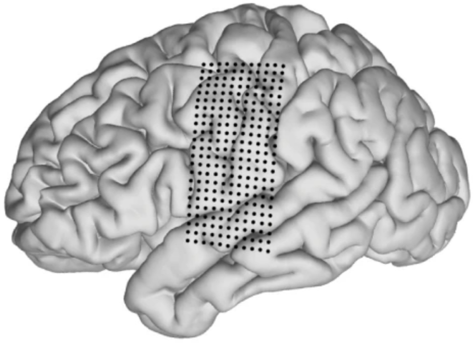
UCSF BRAVO trial:

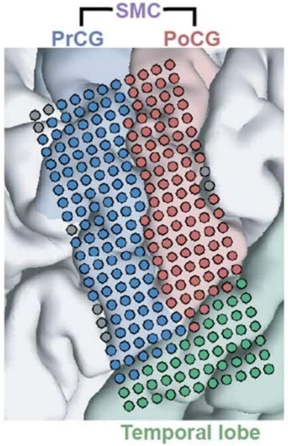
Demonstration in our first clinical trial participant (BRAVO1) was limited to 50 words text communication

A multimodal ECoG speech BCI



Metzger*, Littlejohn*, Silva*, Moses*, Seaton* et al. 2023, *Nature*





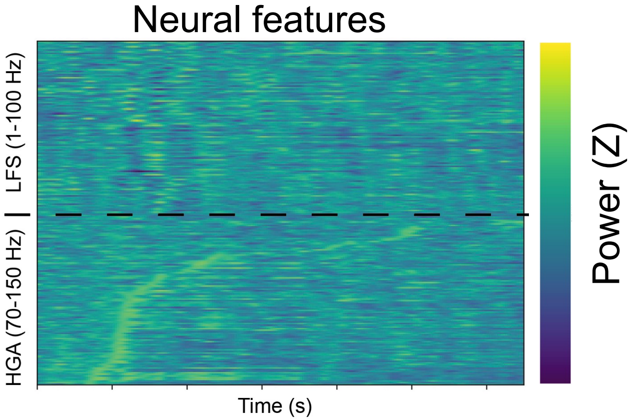
Coverage of the sensorimotor cortex primarily

Sentence sets used for training and testing decoders

- 1024-word-General
  - Over 1,000 English words which can cover over 85% of conversational English
  - Testing was always on unseen sentences

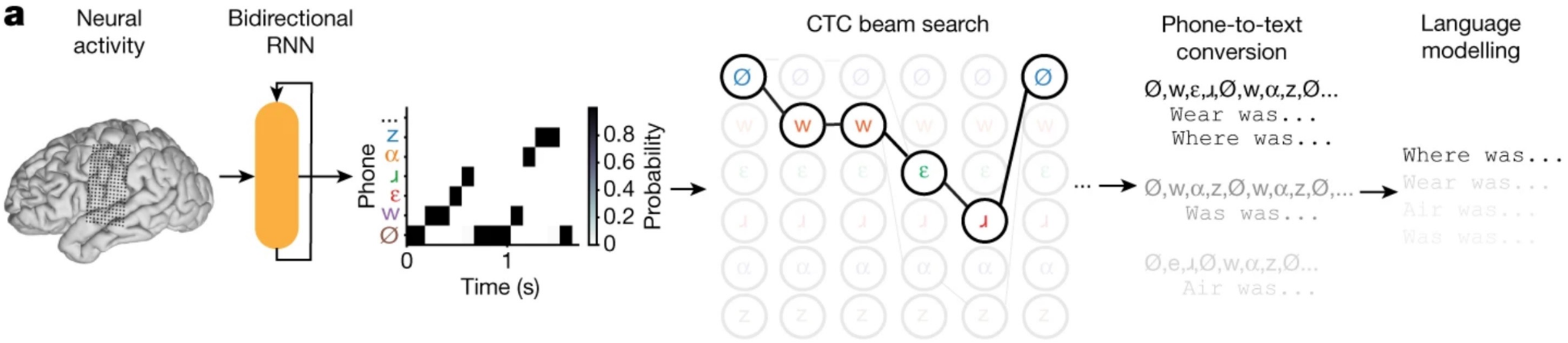
- 529-phrase-AAC
  - 529 phrases relevant for daily life and caregiving

- 50-phrase-AAC
  - A subset of 50 phrases from the prior set

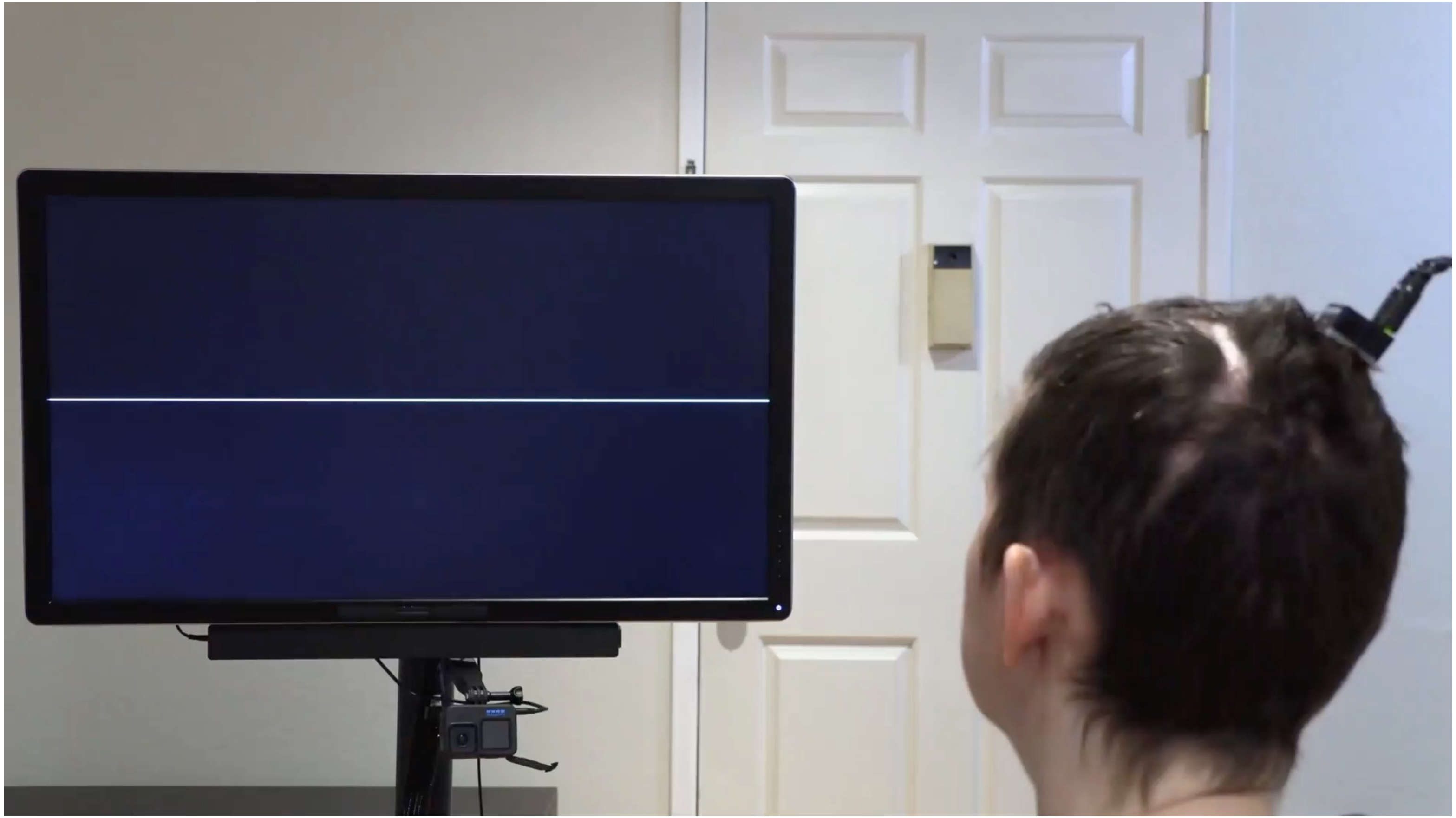


UCSF Chang Lab

A CTC model to decode phonemes and words



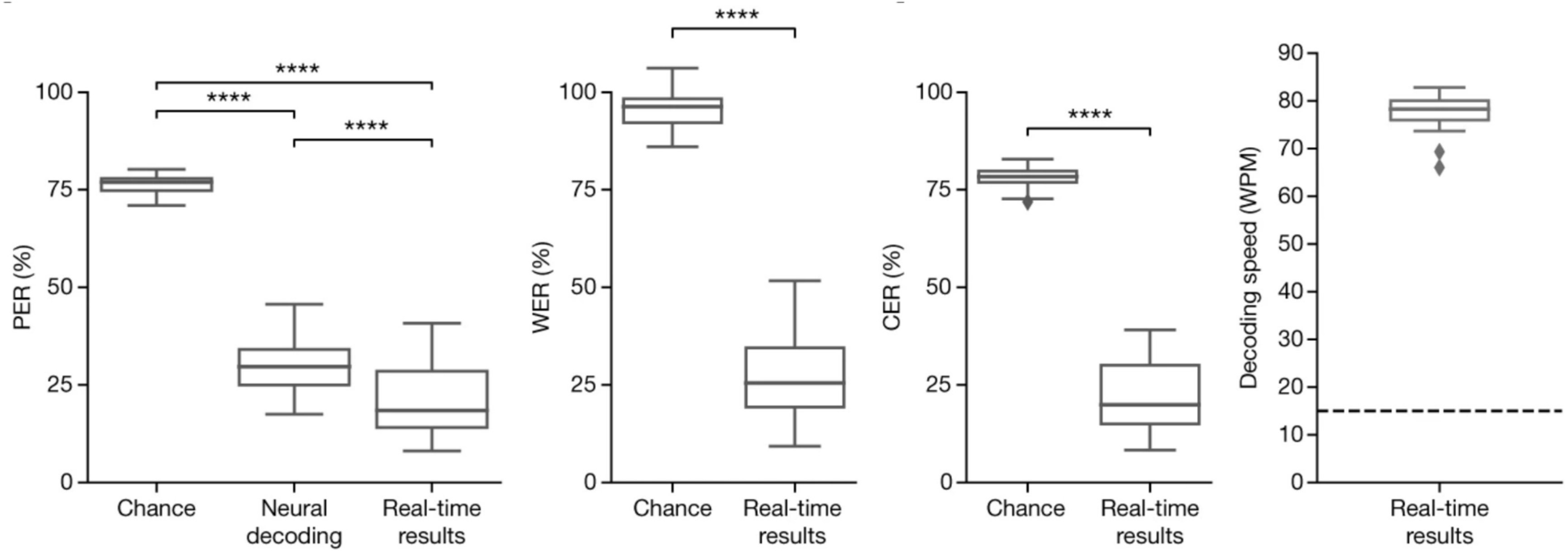
- Recurrent neural network consumes neural activity (HGA: 70-150 Hz, LFS: 1-100 Hz)
- RNN outputs the probability of each phoneme at down-sampled timesteps (emissions)
- Beam search: finds most likely sequence of phonemes and words based on the emissions



High-performance text decoding

Target sentence	Decoded sentence	Word error rate (%)
You should have let me do the talking	You should have let me do the talking	0
I think I need a little air	I think I need a little air	0
Do you want to get some coffee	Do you want to get some coffee	0
What do you want from us	What do you want for us	17
You have no right to keep us here	You have no right to be out here	25
Do you mind me talking about your stuff	Do you make it out to yourself	75

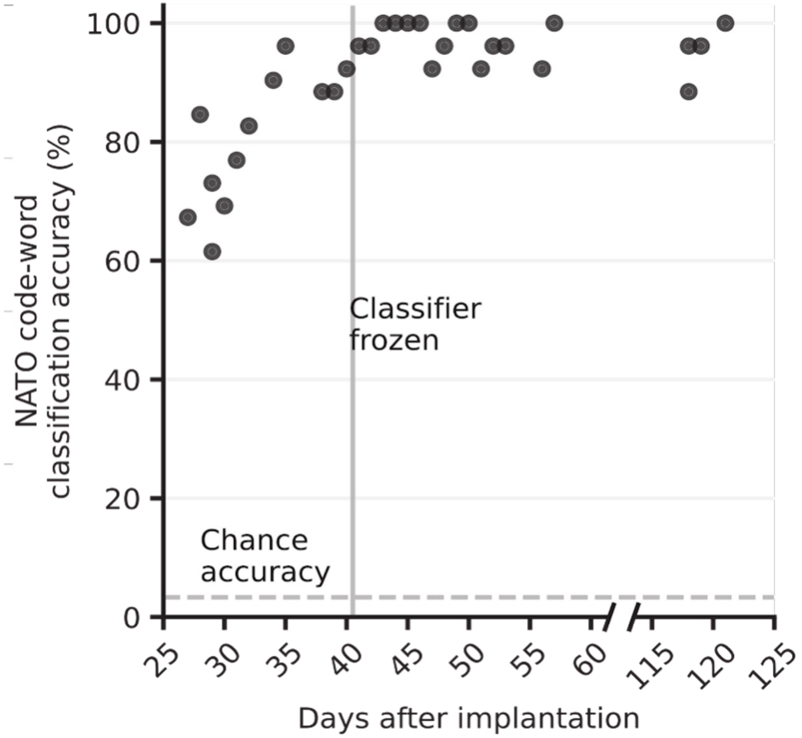
High-performance text decoding with 1024-word-General set



- ~5x higher decoding rate, 20x larger vocabulary, ~ the same WER as prior demonstration

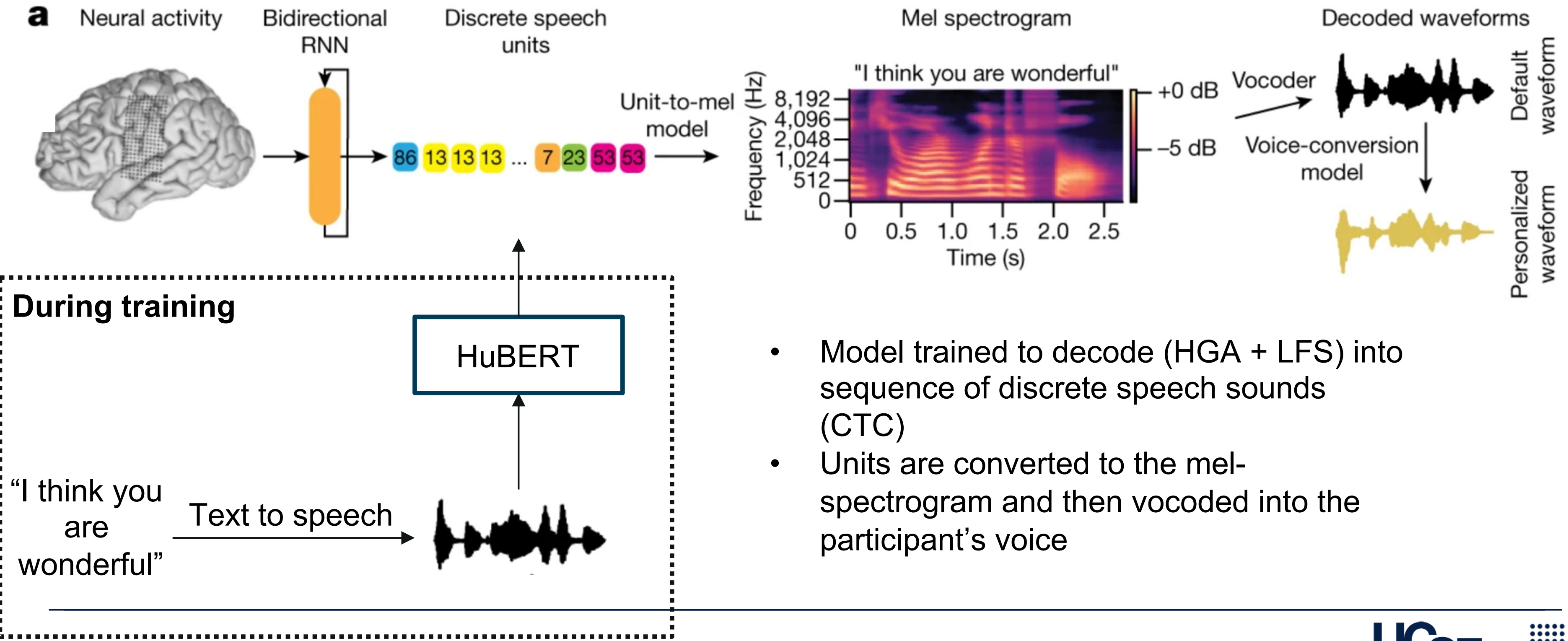
Stable NATO code-word classification

Stable classification accuracy without re-training for ~80 days on 26 NATO code-words



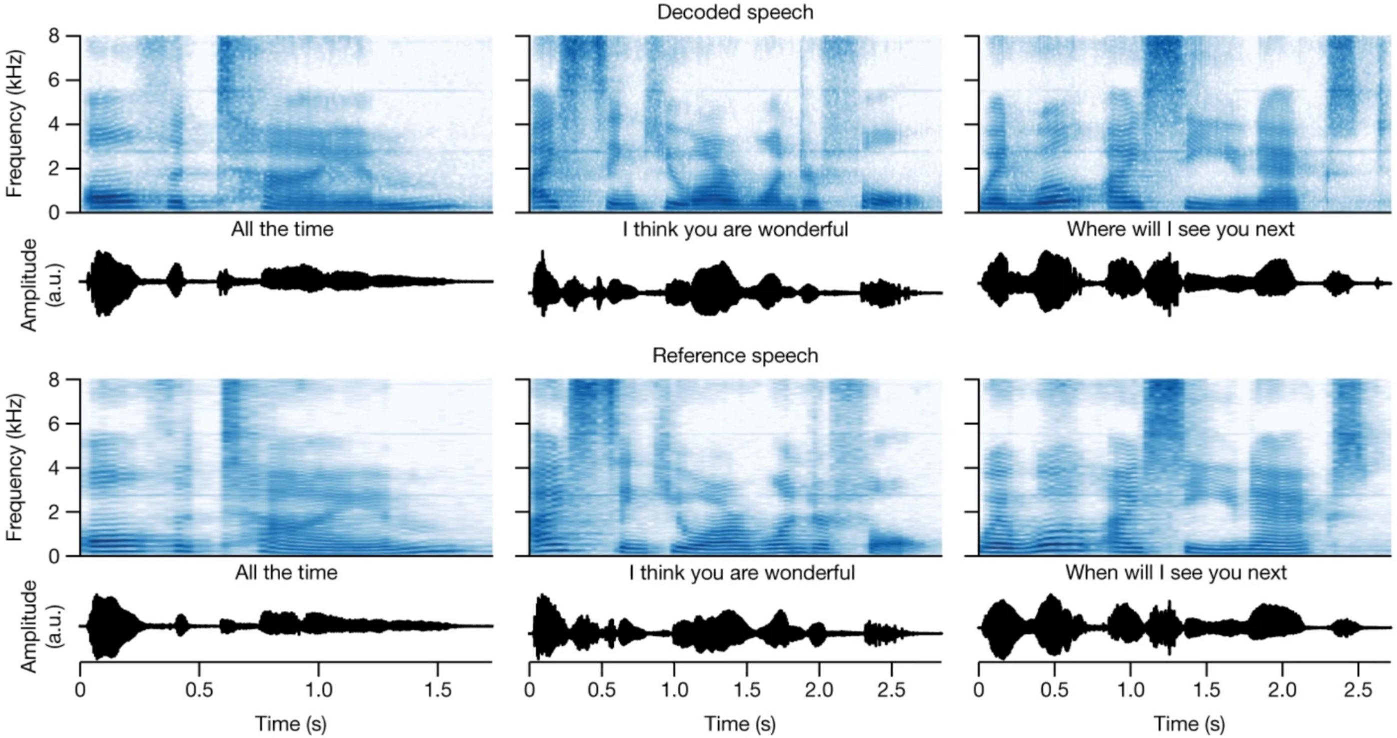
alpha (1)
bravo
charlie
delta
.
.
.
zulu (26)

Intelligible speech synthesis

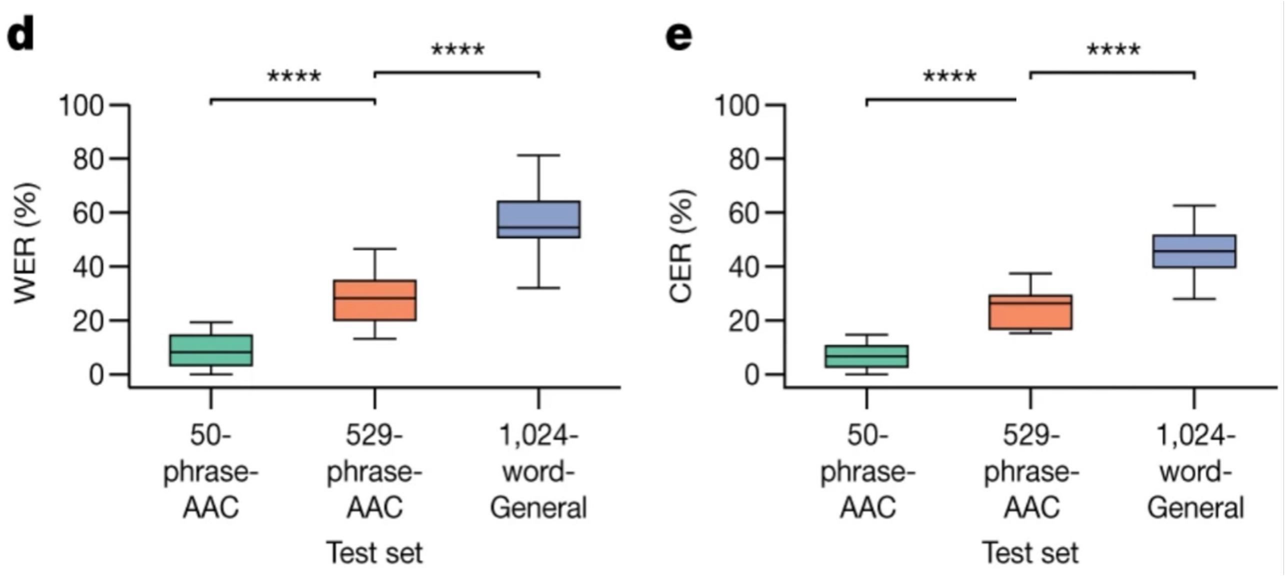


- Model trained to decode (HGA + LFS) into sequence of discrete speech sounds (CTC)
- Units are converted to the mel-spectrogram and then vocoded into the participant's voice

Intelligible speech synthesis

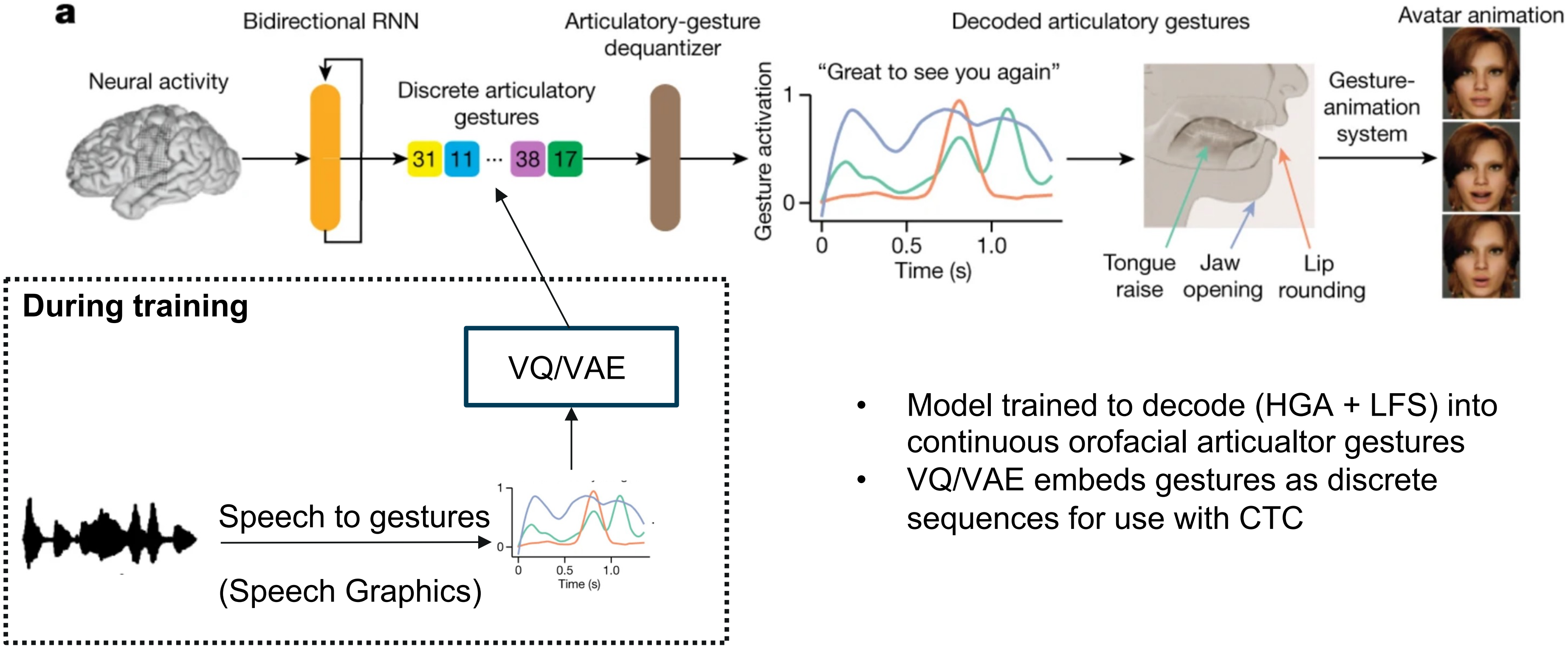


Intelligible speech synthesis



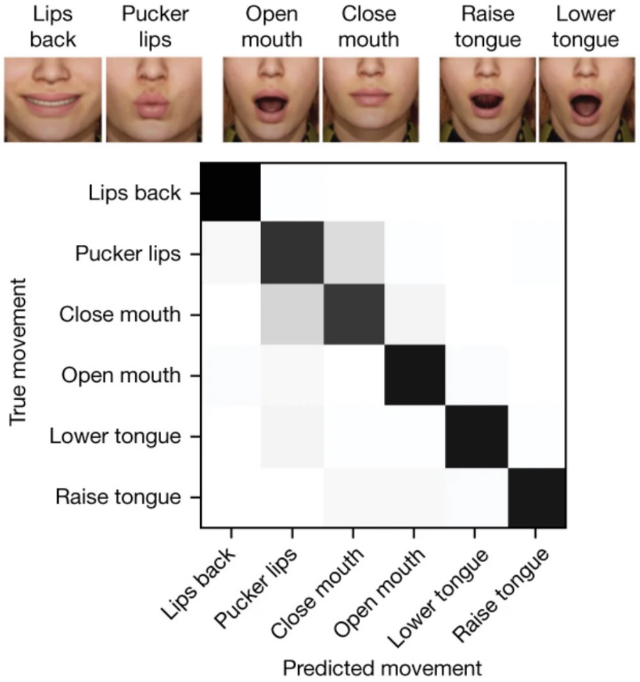
- Volunteers listened to the synthesized audio and transcribed what they heard
- Transcriptions were used to compute WERs and CERs

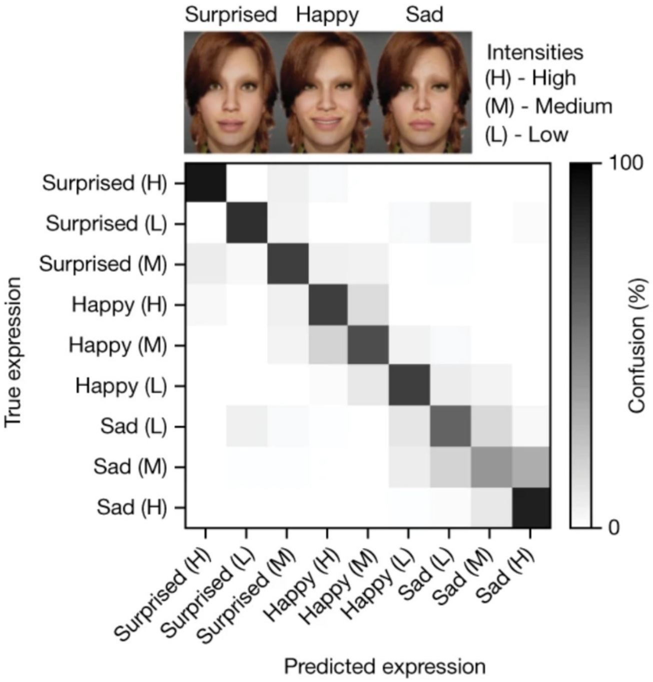
Facial-avatar control



- Model trained to decode (HGA + LFS) into continuous orofacial articualtor gestures
- VQ/VAE embeds gestures as discrete sequences for use with CTC

Facial-avatar control (expressions)



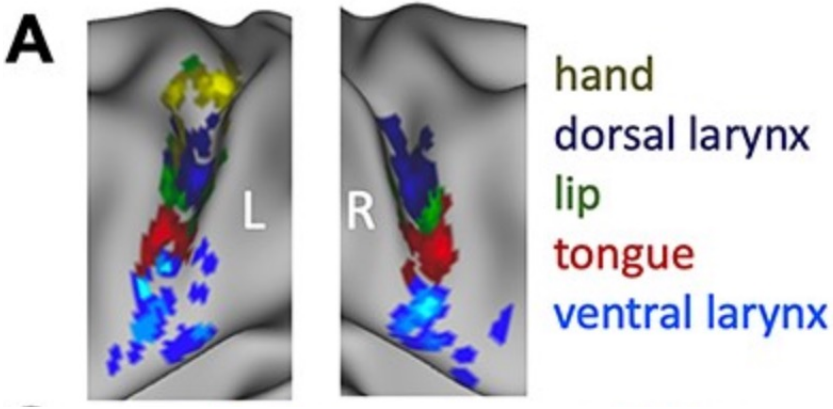


Demonstration of all three modalities

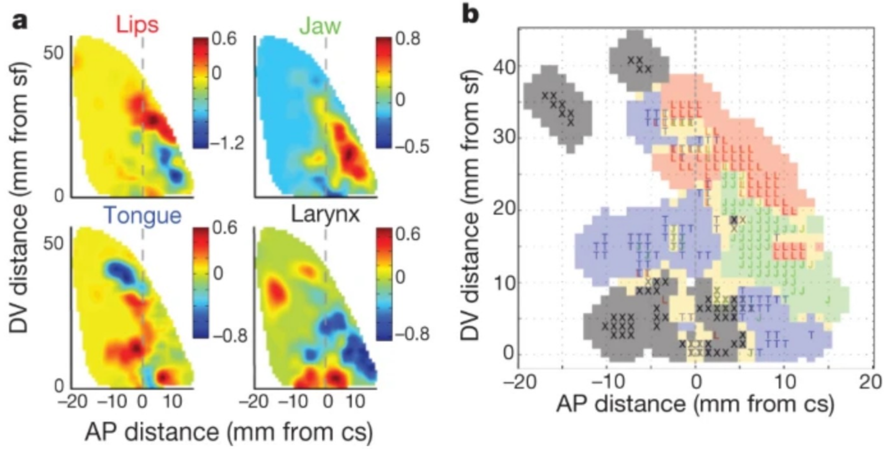
What neural signals are our models picking up on?

To what extent do fine-grain articulatory representations persist in people with paralysis?

- In healthy speakers, speech articulatory representations are arranged somatotopically

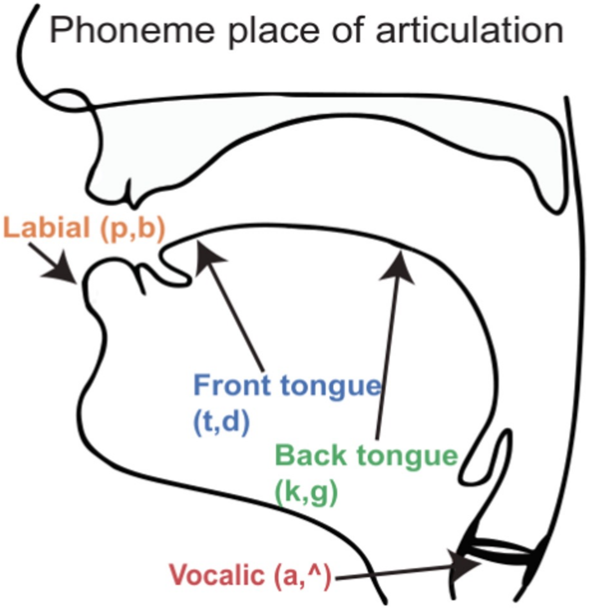


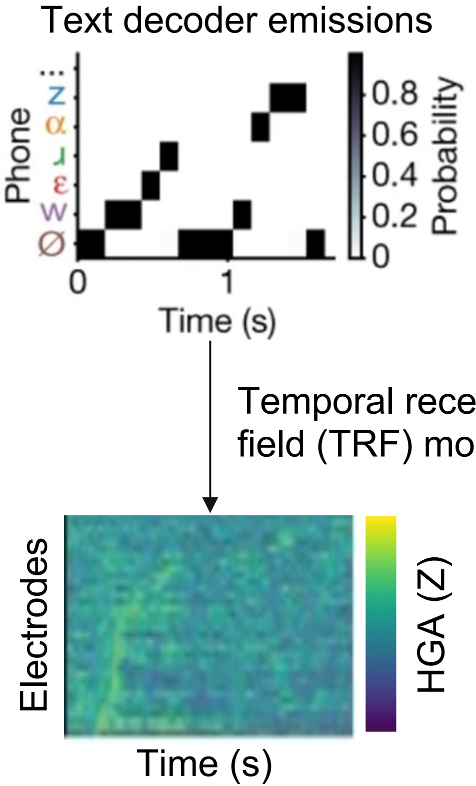
Eichert et al. 2020 *Cerb. Cortex*



Bouchard et al. 2013 *Nature*

Phonemes can be grouped based on their place of articulation



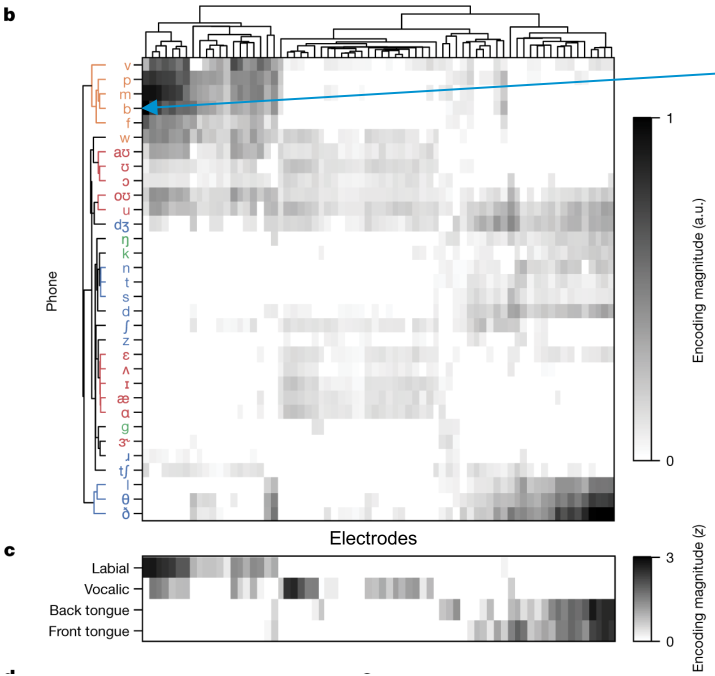


Temporal receptive field (TRF) model

TRF:

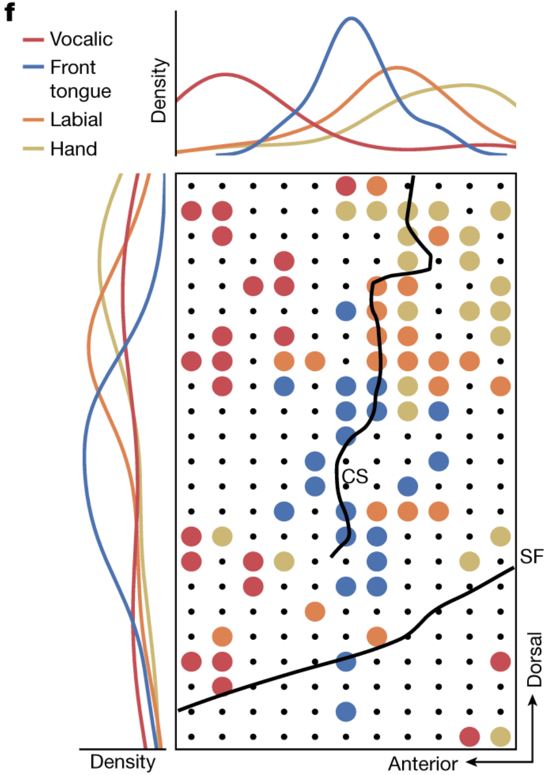
- R^2, how well can HGA be predicted from phonemes
- Coefficients, how much does each phoneme contribute to an electrode's HGA

Somatotopy persists after 18 years of paralysis



Selectivity of electrode 1 to ‘b’ phoneme

- Electrodes have selectivity for groups of phonemes articulated at some location
- Somatotopic organization parallels that of healthy speakers



UCSF Chang Lab

Comments from participant

- Hearing a voice similar to your own is emotional. Being able to have the ability to speak aloud is very important.
- My moonshot was to become a counselor and use the system to talk to my clients. I think the avatar would make them more at ease.
- Please make the device wireless!

Metzger, S.L., Littlejohn, K.T., Silva, A.B. et al. A high-performance neuroprosthesis for speech decoding and avatar control. Nature 2023

Summary

- Rapid and high-performance text-decoding
- First intelligible and personalized speech synthesis with someone who cannot speak
- First demonstration of real-time avatar-control
- Detailed articulatory representations persist after paralysis

Acknowledgments

- **Our great participants + their caregivers & family**
- **Members of the Ganguly lab**
- **The Speech Graphics team**
- **UCSF and Berkeley support staff**
- **Funding sources**
  - National Institutes of Health
  - Joan and Sandy Weill and the Weill Family Foundation
  - The Bill and Susan Oberndorf Foundation
  - Ron Conway
  - Graham and Christina Spencer
  - The William K. Bowes, Jr. Foundation
  - Rose Hills and Noyce Foundations
  - National Institute of General Medical Sciences
  - National Science Foundation

Edward Chang

Kaylo Littlejohn

Sean Metzger

Margaret Seaton

David Moses

Ran Wang

Max Dougherty

Jessie Liu

Peter Wu

Michael Berger

Inga
Zhuravleva

Adelyn
Tu-Chan

Karunesh
Ganguly

Gopala
Anumanchipalli

UCSF Chang Lab