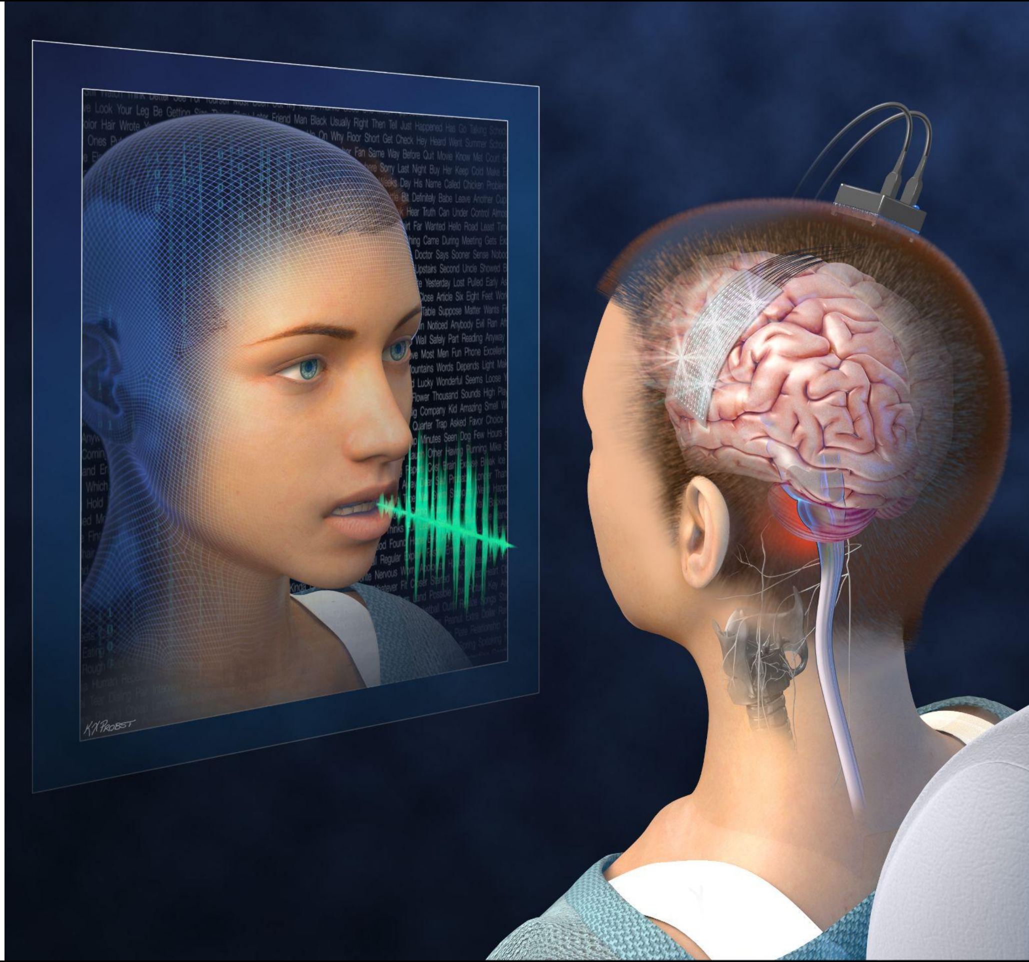
Speech neuroprostheses for restoring naturalistic communication and future prospects

Cheol Jun Cho

UCSF
University of California
San Francisco

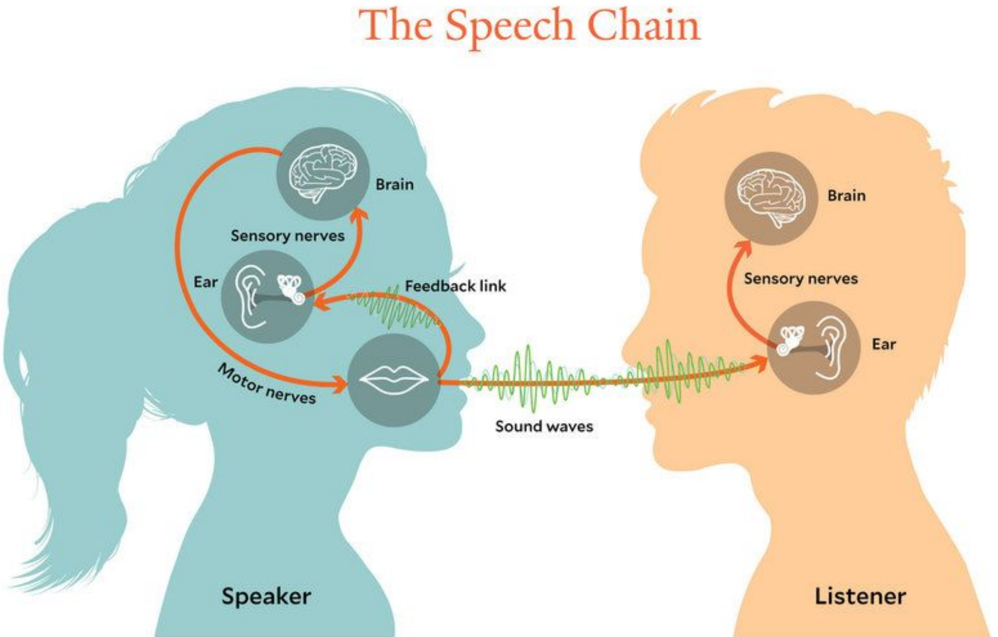
Chang Lab

Berkeley
UNIVERSITY OF CALIFORNIA



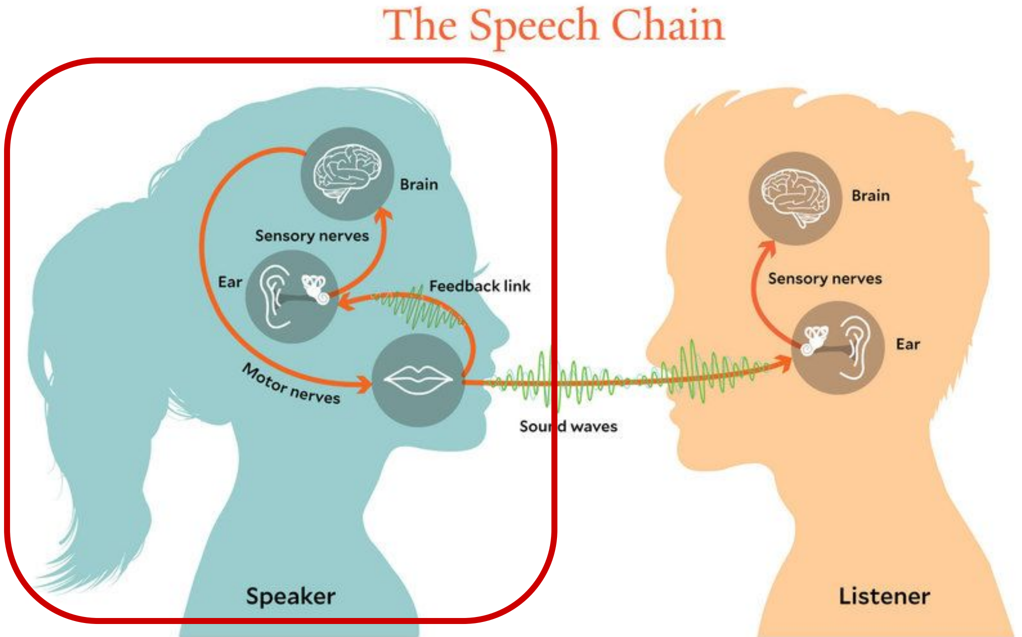
Speech is a natural and effective tool of human communication

- Evolved for more than ***100,000 years.***
- More than ***7,000 languages*** exist around the world.
- On average, individuals speak over ***15,000 words*** per day.
- One of the most complicated cognitive processes of the human.

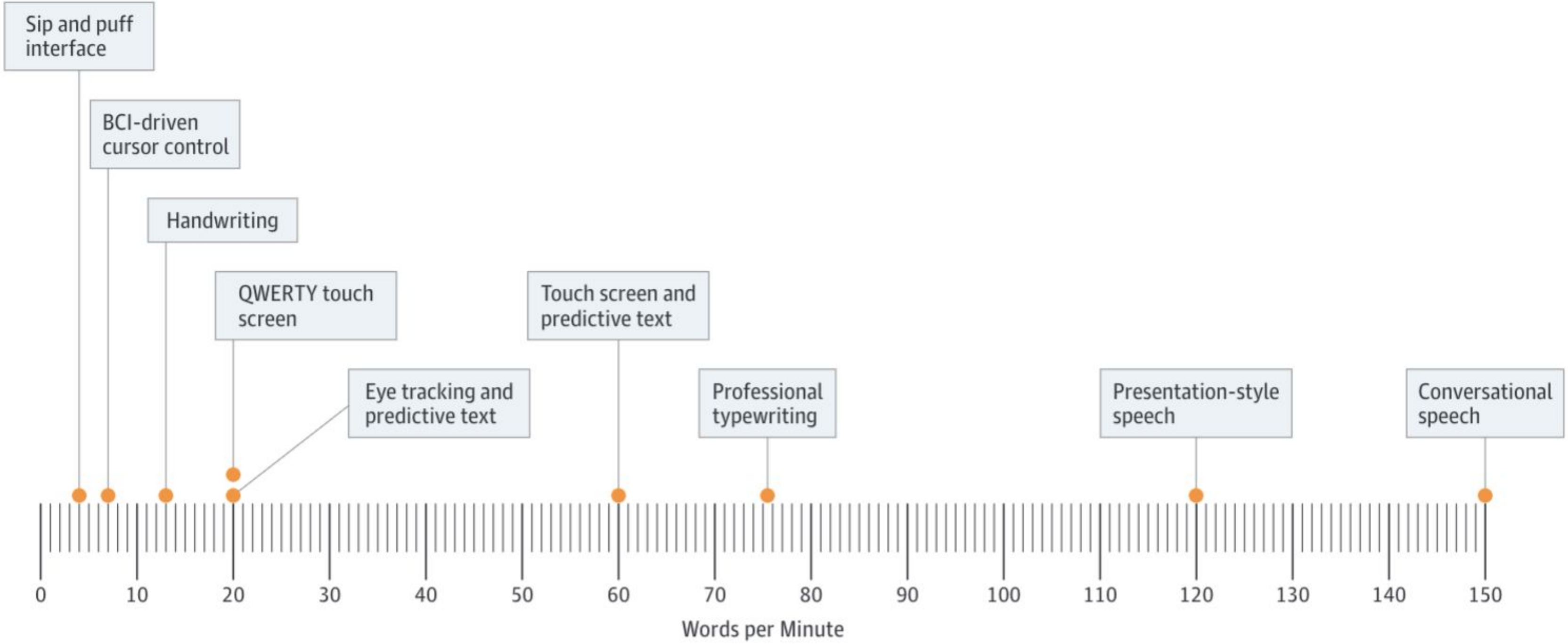


Speech is a natural and effective tool of human communication

- Evolved for more than ***100,000 years.***
- More than ***7,000 languages*** exist around the world.
- On average, individuals speak over ***15,000 words*** per day.
- One of the most complicated cognitive processes of the human.
- ***Impairments of the ability to speak significantly affect quality of life.***

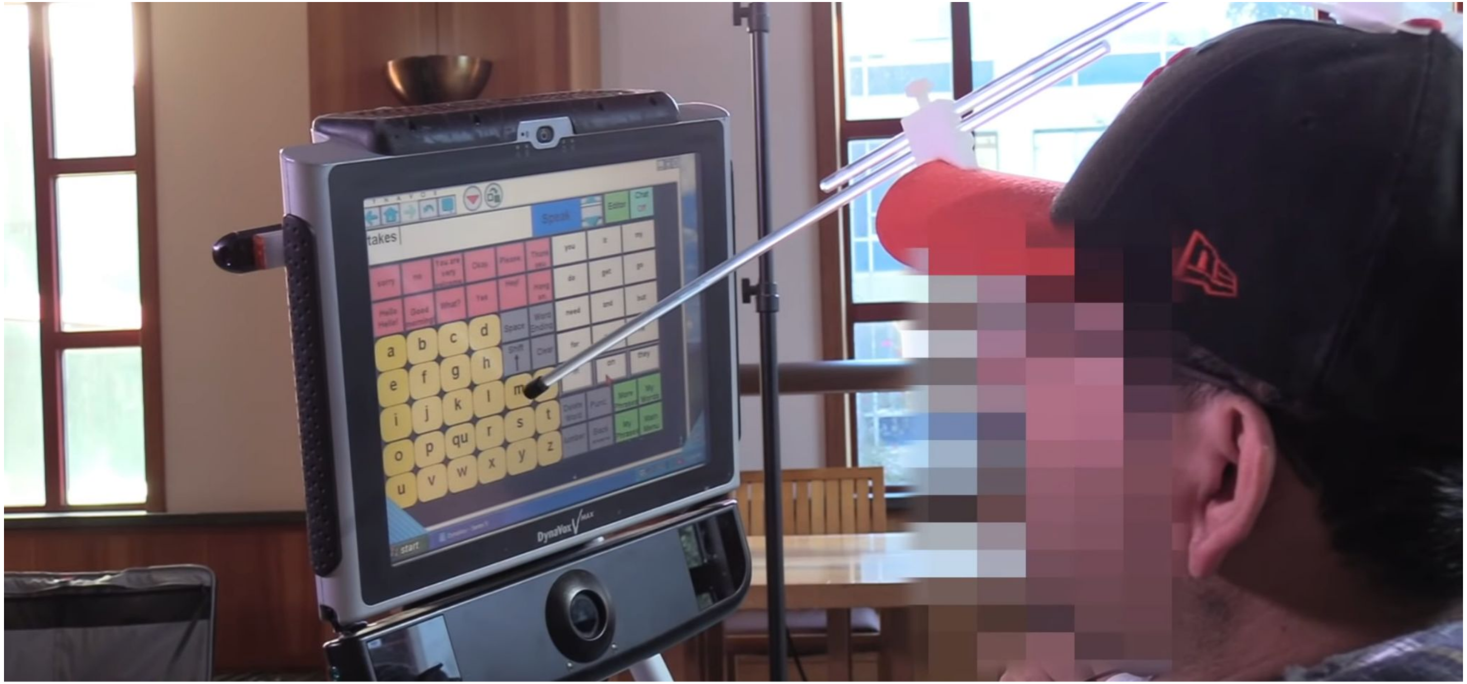


Current assistive communication technology is much slower than speech



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

Standard of care for assistive communication



Operates at a speed of ~5-15 words per minute

A brain-computer interface to decode speech from brain activity could help persons with paralysis

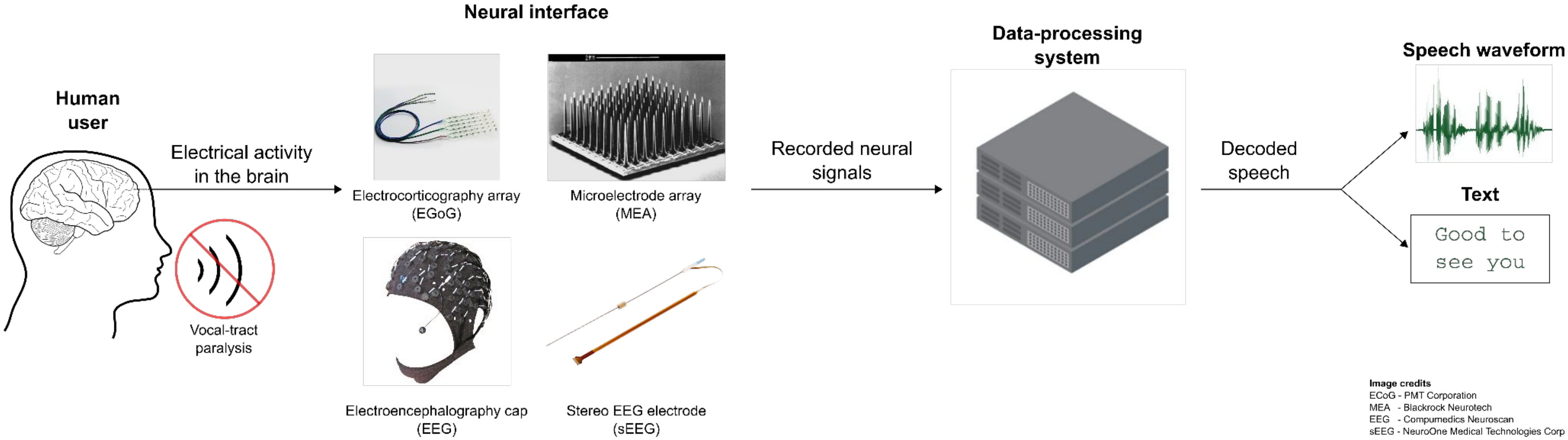
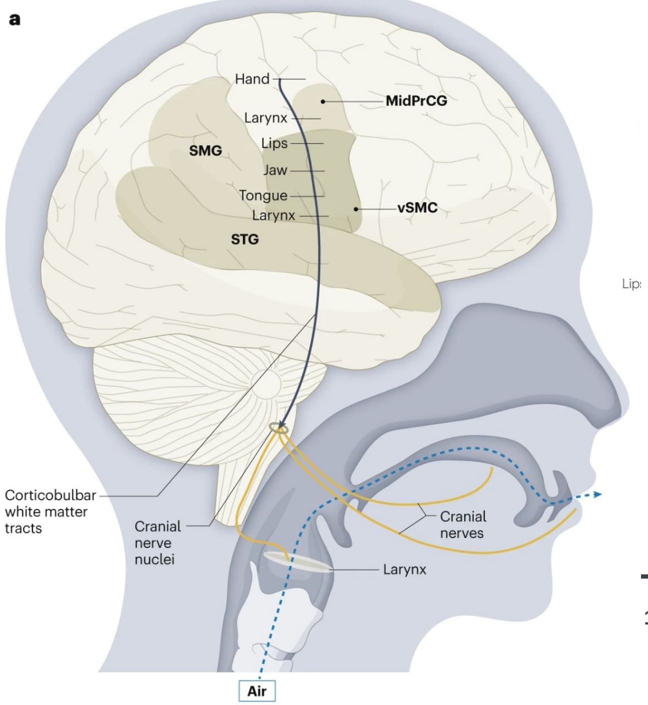
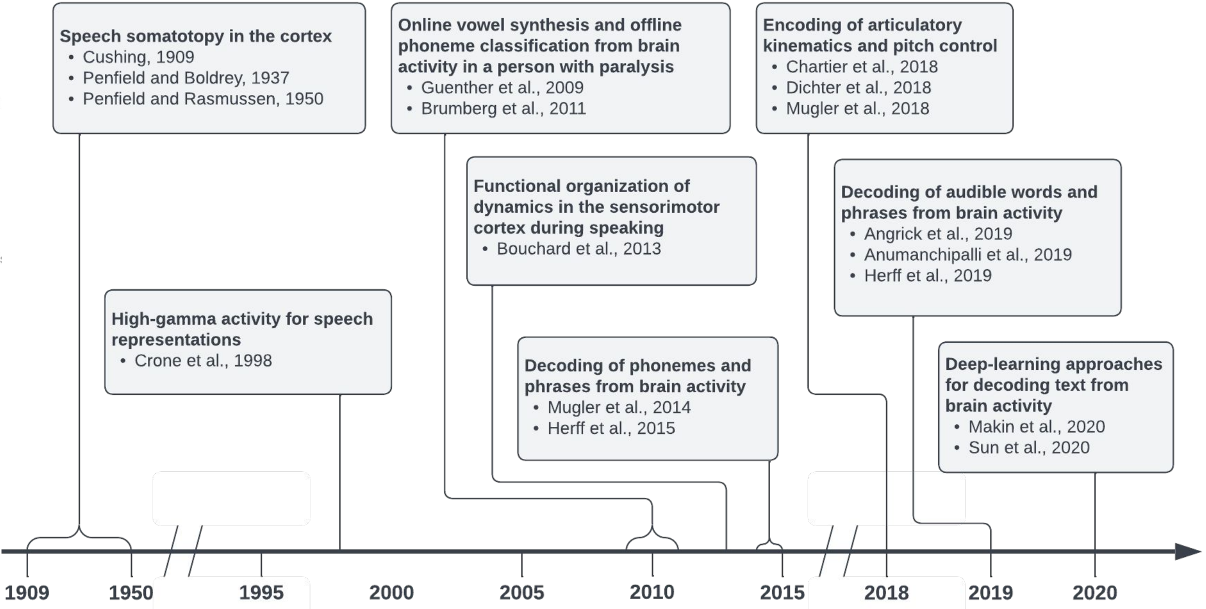

Neural interface
Human user
Electrical activity in the brain
Vocal-tract paralysis
Electrocorticography array (EGoG)
Microelectrode array (MEA)
Electroencephalography cap (EEG)
Stereo EEG electrode (sEEG)
Recorded neural signals
Data-processing system
Decoded speech
Speech waveform
Text
Good to see you

Image credits
ECoG - PMT Corporation
MEA - Blackrock Neurotech
EEG - Compumedics Neuroscan
sEEG - NeuroOne Medical Technologies Corp

Over a century of research characterizing speech and motor control in the brain





Speech somatotopy in the cortex
- Cushing, 1909
- Penfield and Boldrey, 1937
- Penfield and Rasmussen, 1950

High-gamma activity for speech representations
- Crone et al., 1998

Online vowel synthesis and offline phoneme classification from brain activity in a person with paralysis
- Guenther et al., 2009
- Brumberg et al., 2011

Functional organization of dynamics in the sensorimotor cortex during speaking
- Bouchard et al., 2013

Decoding of phonemes and phrases from brain activity
- Mugler et al., 2014
- Herff et al., 2015

Encoding of articulatory kinematics and pitch control
- Chartier et al., 2018
- Dichter et al., 2018
- Mugler et al., 2018

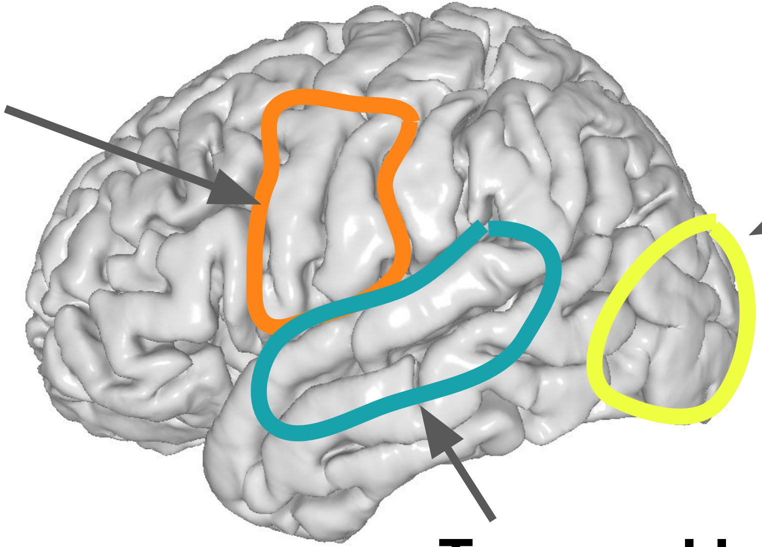
Decoding of audible words and phrases from brain activity
- Angrick et al., 2019
- Anumanchipalli et al., 2019
- Herff et al., 2019

Deep-learning approaches for decoding text from brain activity
- Makin et al., 2020
- Sun et al., 2020

(More recent work discussed in paper)

Silva et al. The speech neuroprosthesis. Nature Reviews Neuroscience 2024

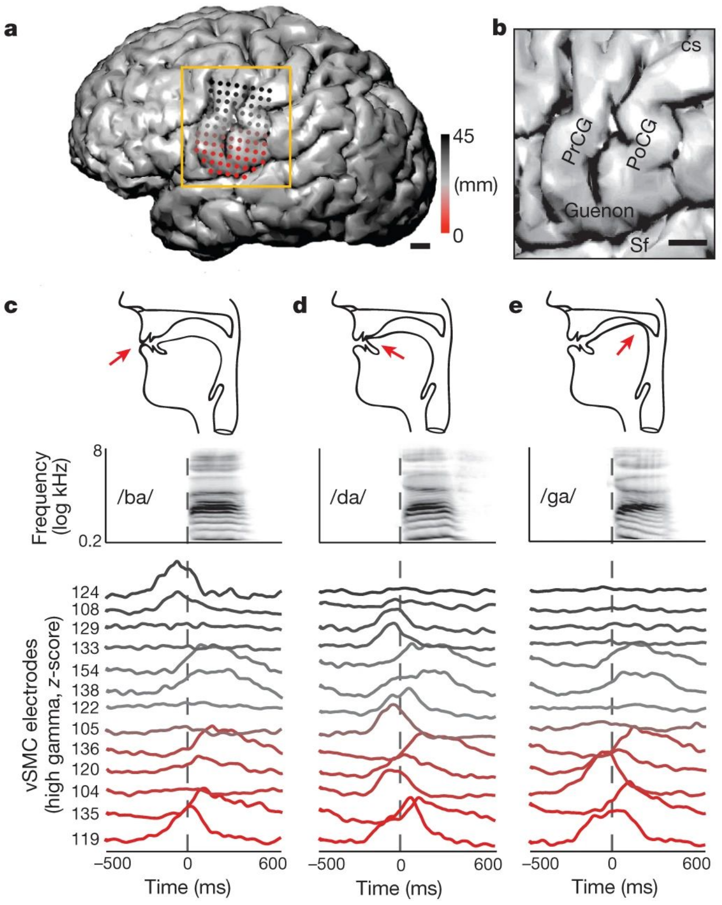
Our brains act as a control center

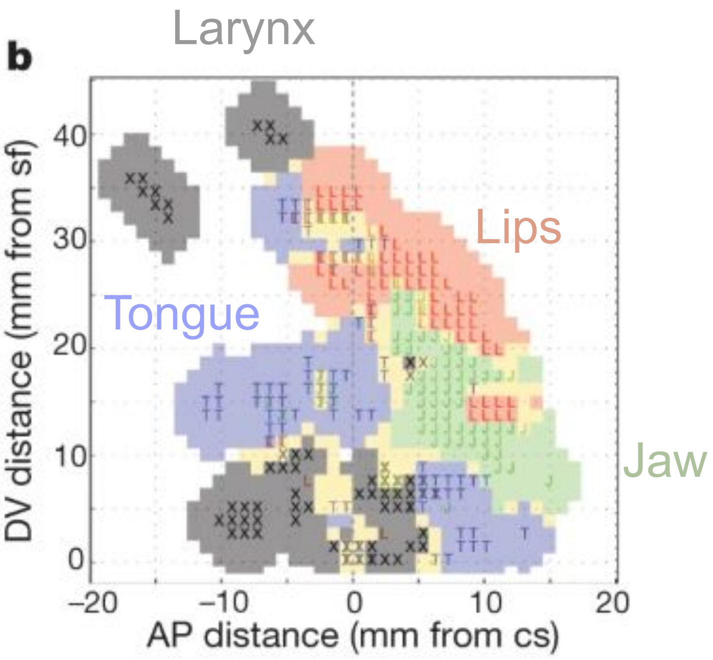


Motor cortex: Controls many of our voluntary movements (arms, mouth, and more)

Visual cortex: Helps interpret what our eyes see

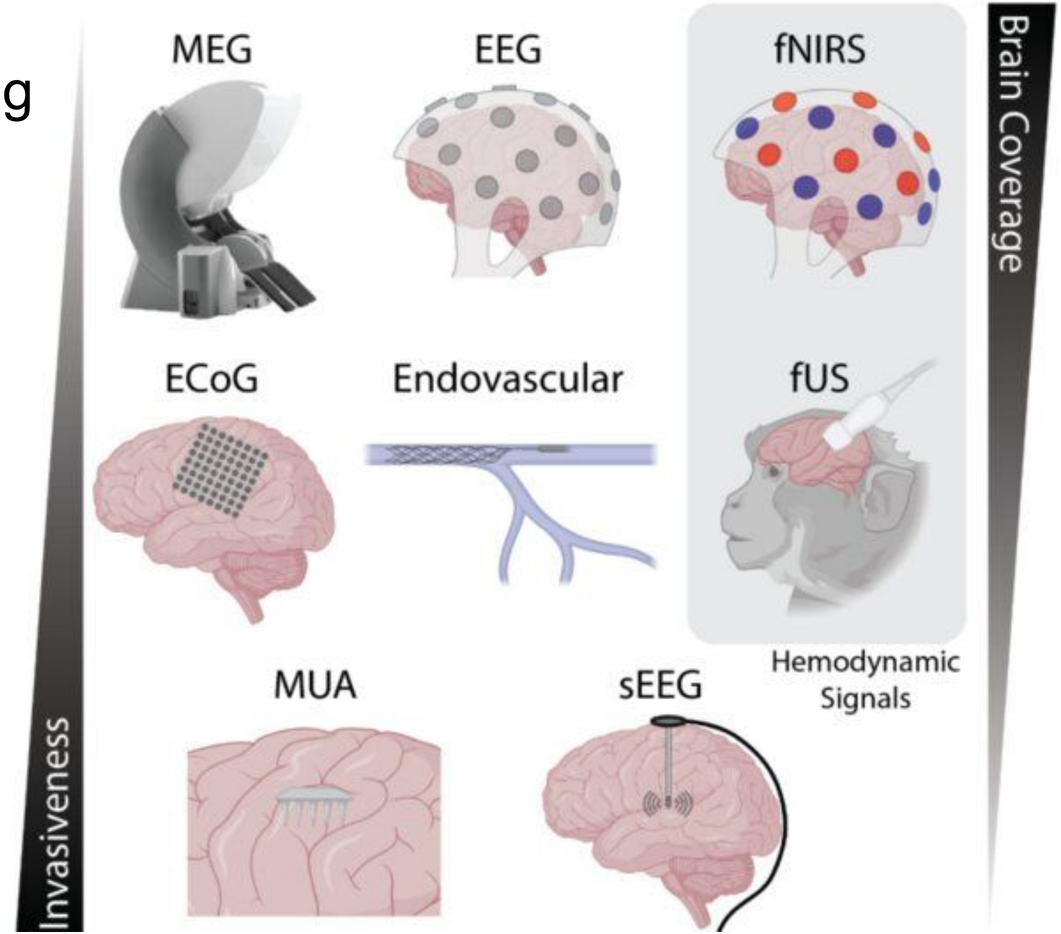
Temporal lobe: Plays a role in understanding what we hear



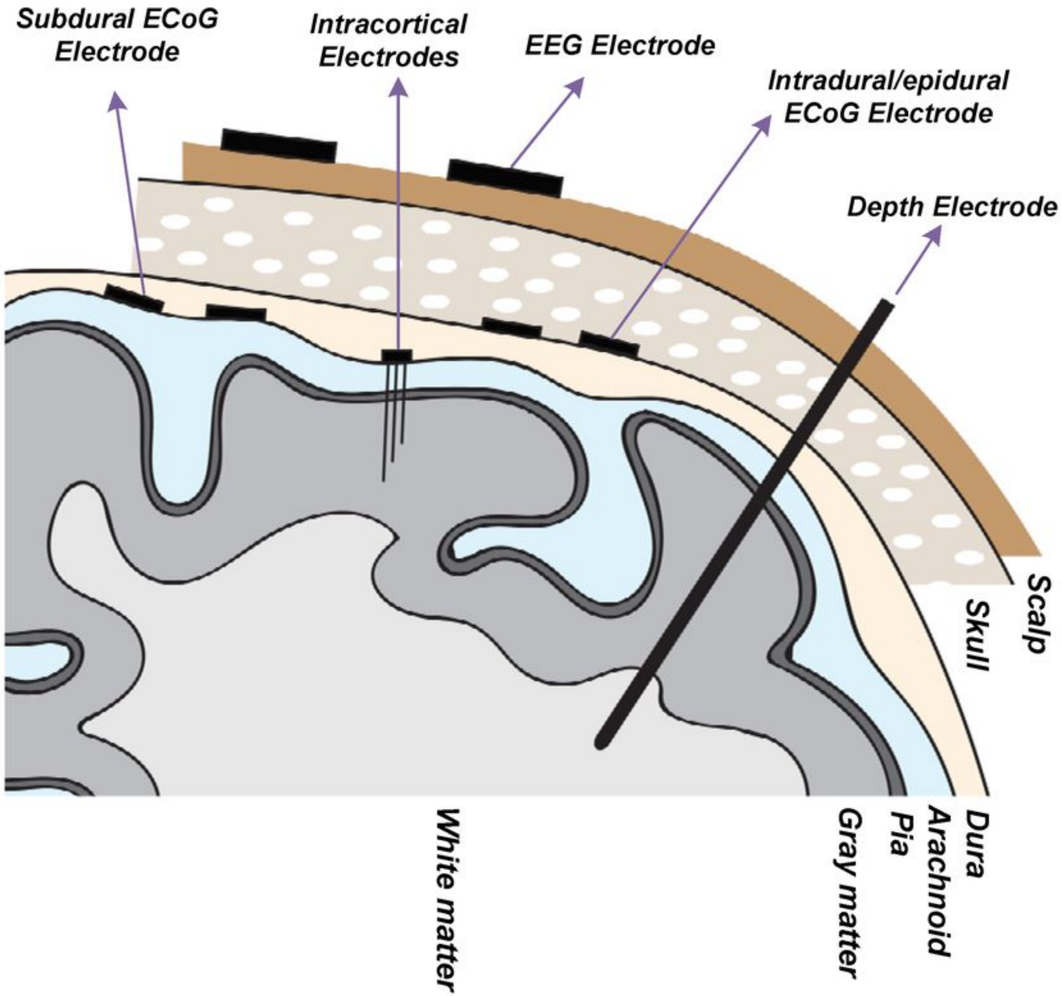


Bouchard et al. (2013)

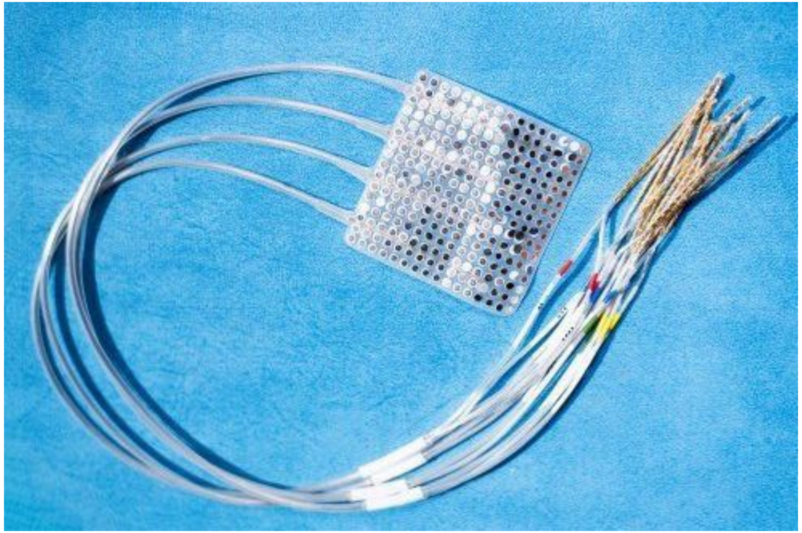
Types of neural recording device

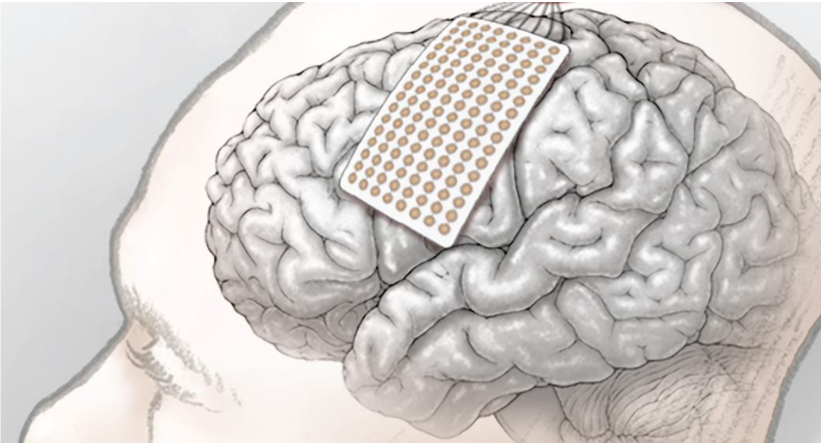


Edelman et al., Non-Invasive Brain-Computer Interfaces: State of the Art and Trends. 2024

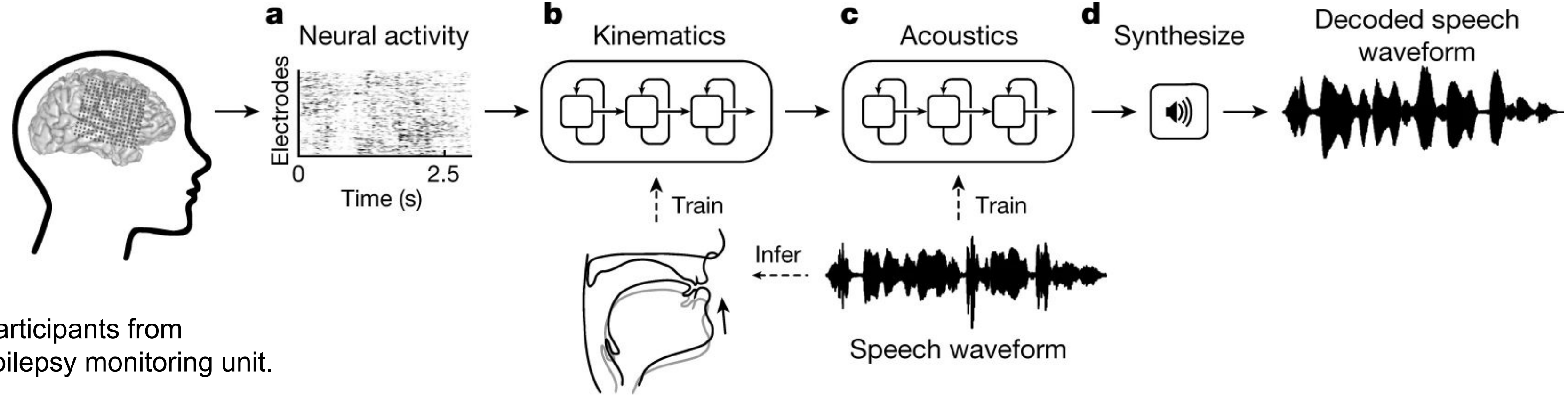

Subdural ECoG Electrode
Intracortical Electrodes
EEG Electrode
Intradural/epidural ECoG Electrode
Depth Electrode
Scalp
Skull
White matter
Gray matter
Pia
Arachnoid
Dura

Electrocorticography (ECoG) to record electrical signals from the brain surface



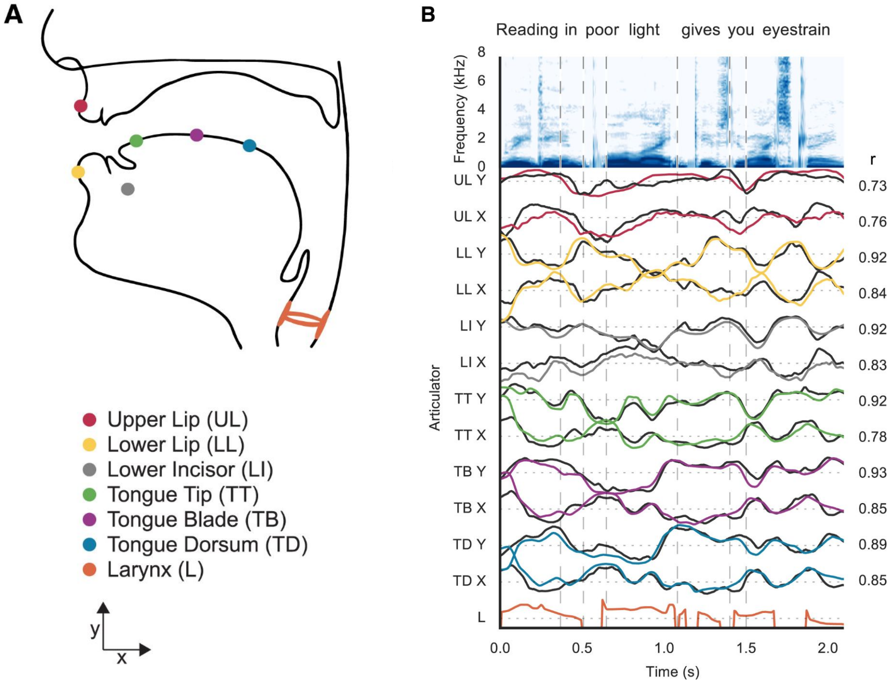


Intelligible and naturalistic speech decoding from brain signals from able speakers

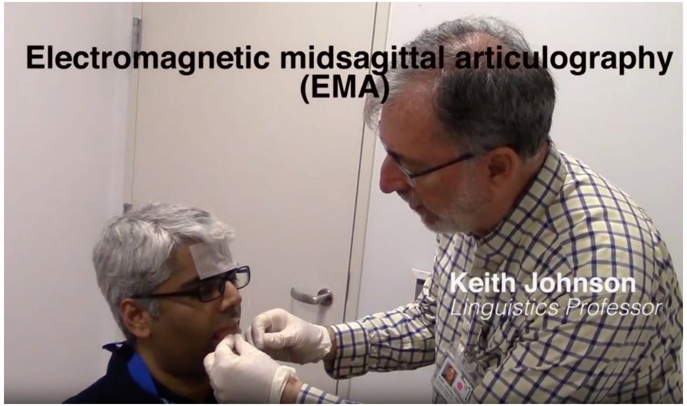


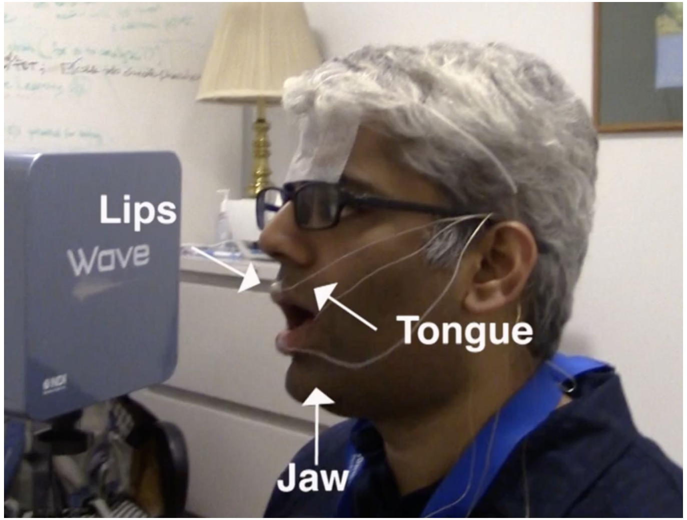
Participants from
epilepsy monitoring unit.

Anumanchipalli, Chartier, and Chang. Speech synthesis from neural decoding of spoken sentences. Nature 2019

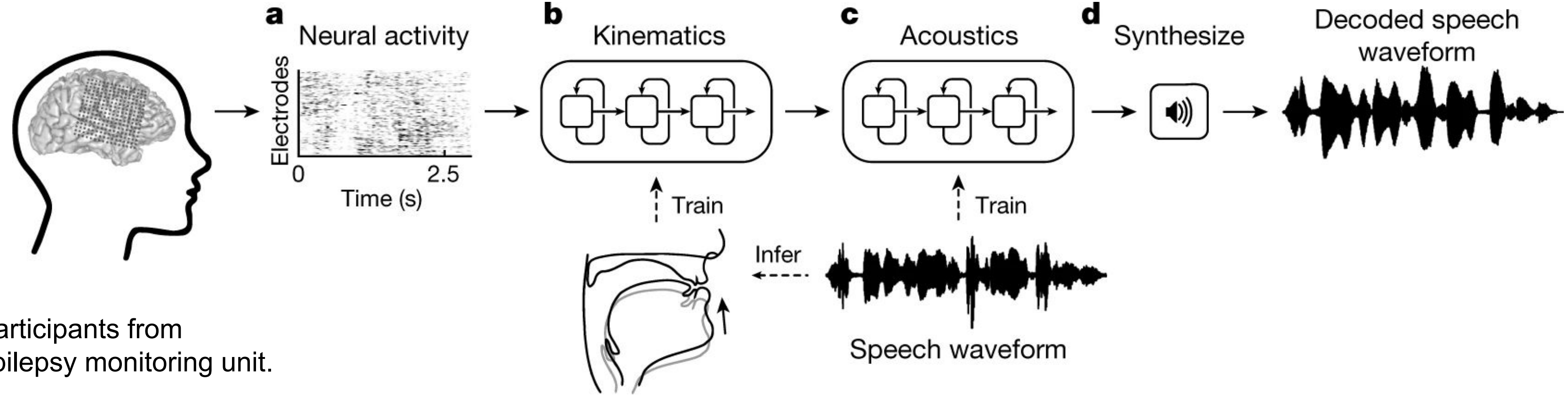


Chartier, Anumanchipalli et al. (2018)





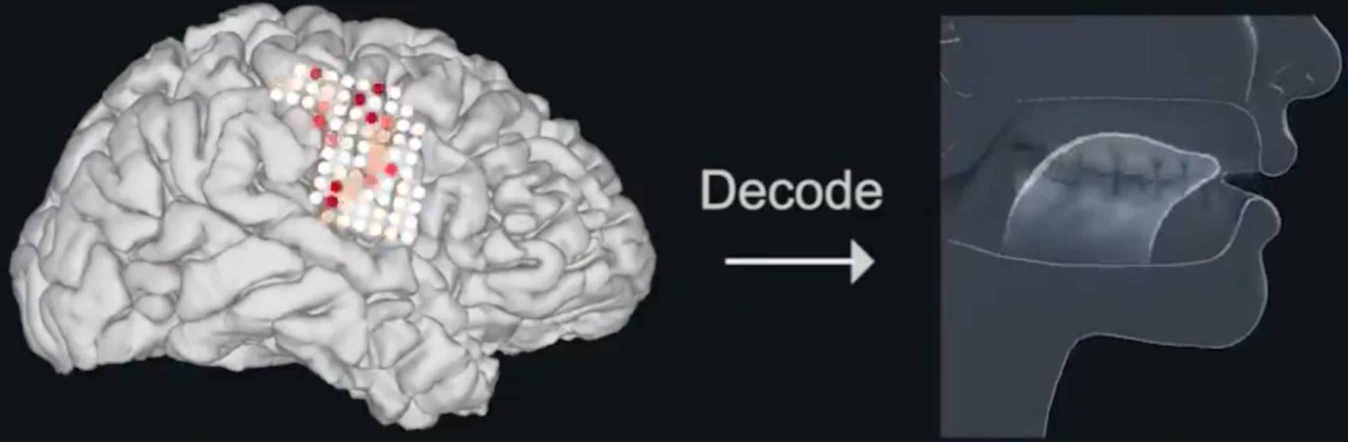
Intelligible and naturalistic speech decoding from brain signals from able speakers



Participants from
epilepsy monitoring unit.

Anumanchipalli, Chartier, and Chang. Speech synthesis from neural decoding of spoken sentences. Nature 2019

Speech synthesized from brain activity



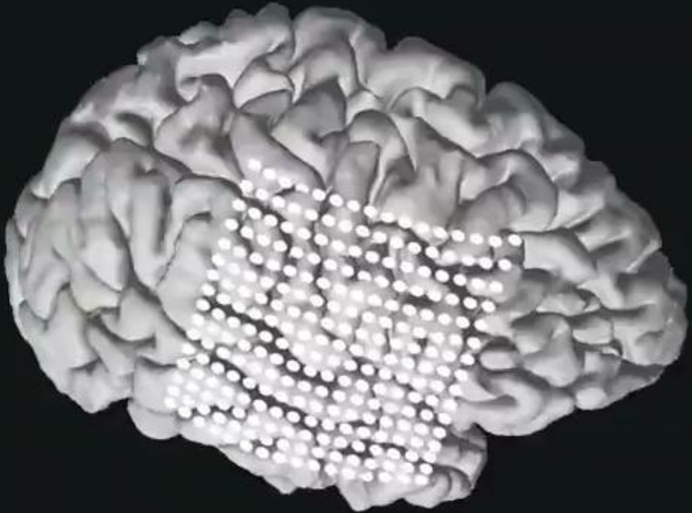
Decode

Synthesize

"The proof you are seeking is not available in books."

UCSF

Speech synthesized from brain activity

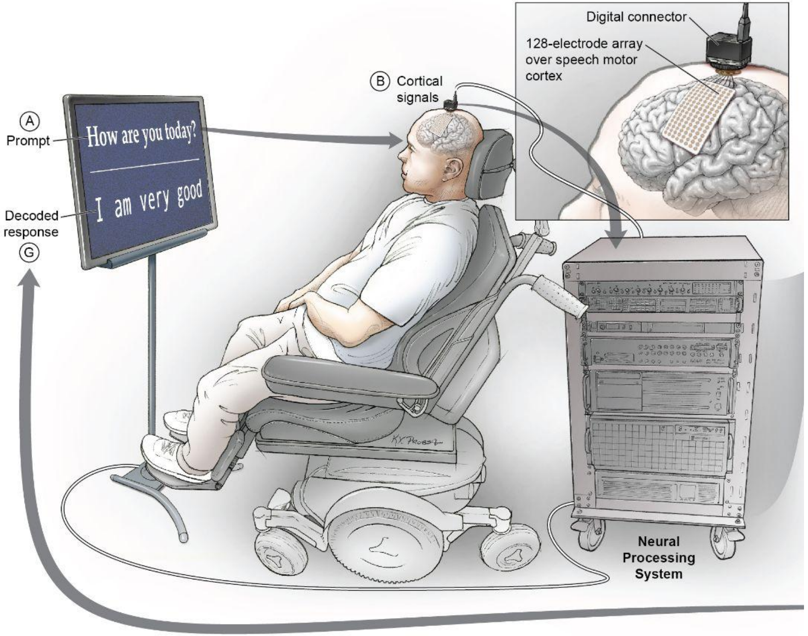


UCSF

“Although I cannot move and I have to speak through a computer, **in my mind I am free.”**

Hawking

Restoring naturalistic communication ability of patients with severe paralysis

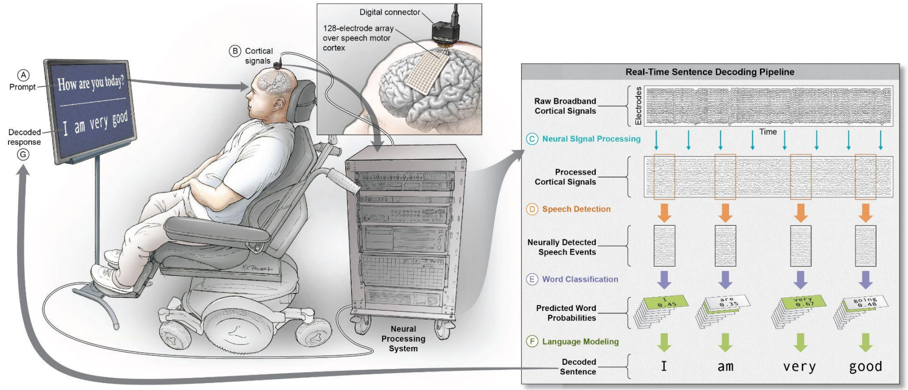


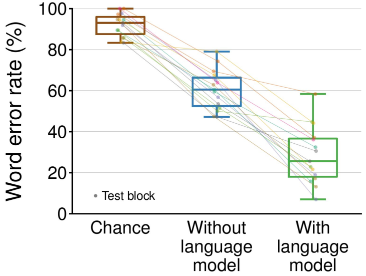
BRAVO: **B**CI **R**estoration of **A**rm and **Vo**ice

- Clinical trial for long-term communication and movement restoration

Moses, Metzger, Liu, et al. Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria. N Engl J Med. 2021

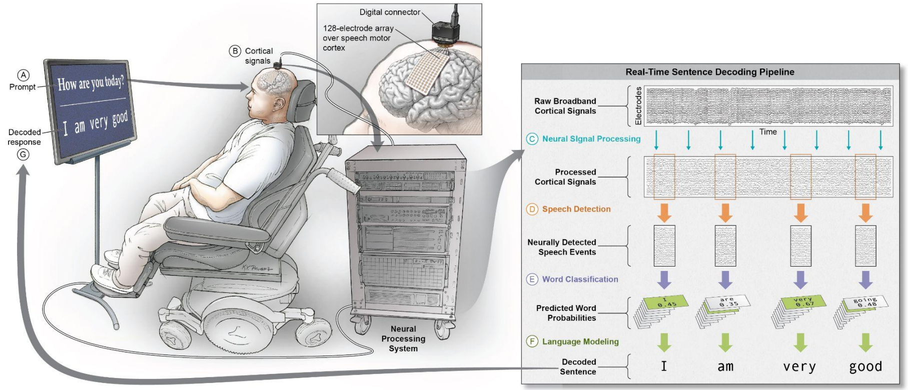
Word detection and classification (50 words)

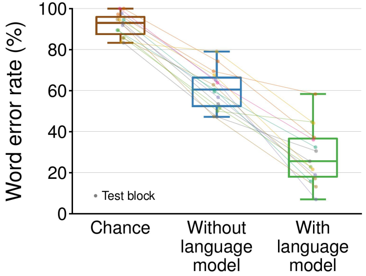




Moses, Metzger, Liu, et al. (2021)

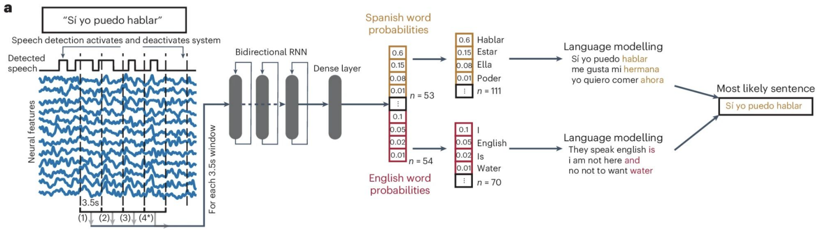
Word detection and classification (50 words)





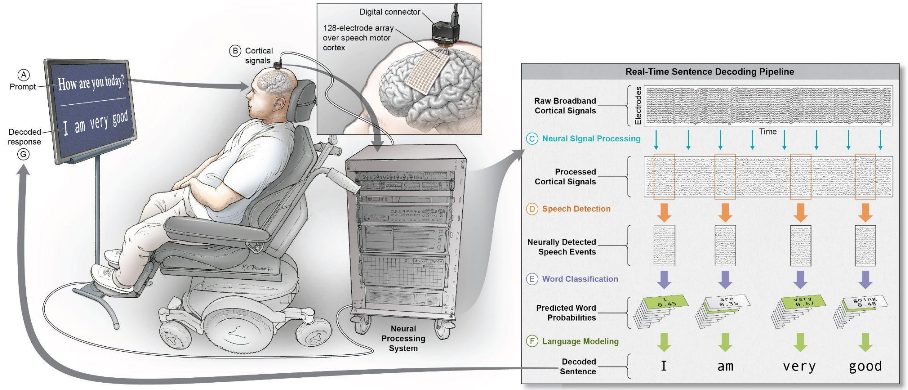
Moses, Metzger, Liu, et al. (2021)

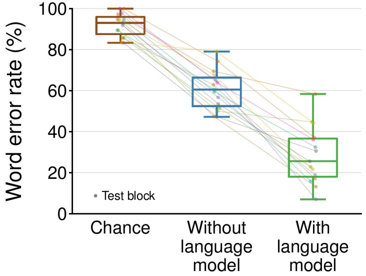
Multilingual: English + Spanish



Silva et al. (2024)

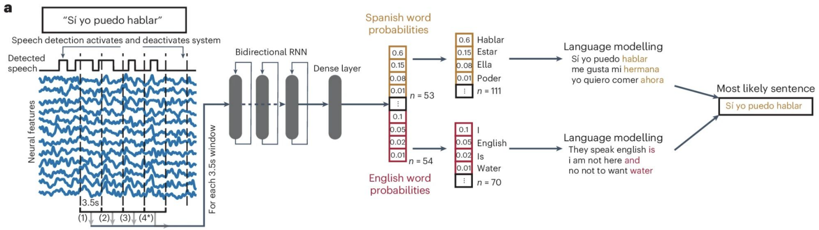
Word detection and classification (50 words)





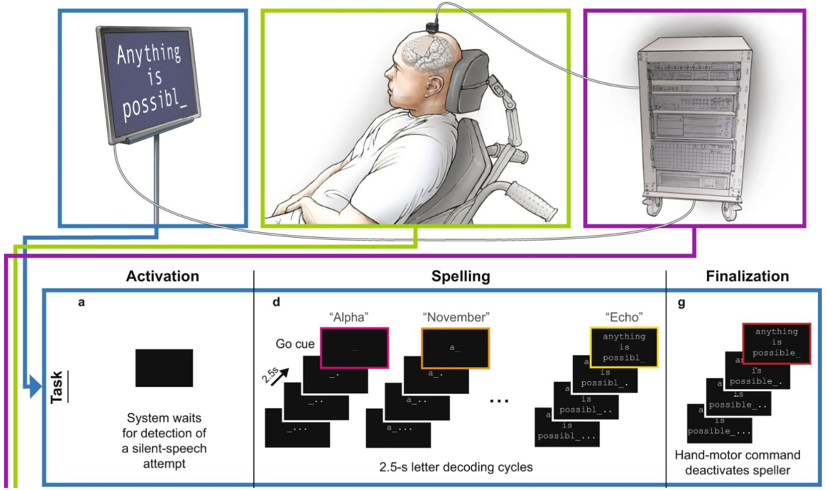
Moses, Metzger, Liu, et al. (2021)

Multilingual: English + Spanish



Silva et al. (2024)

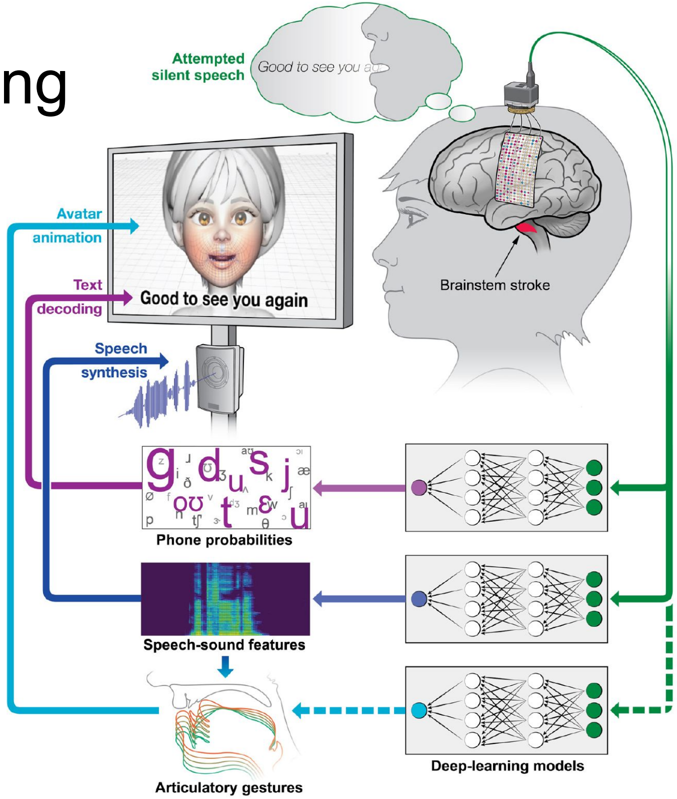
Speller: character decoding



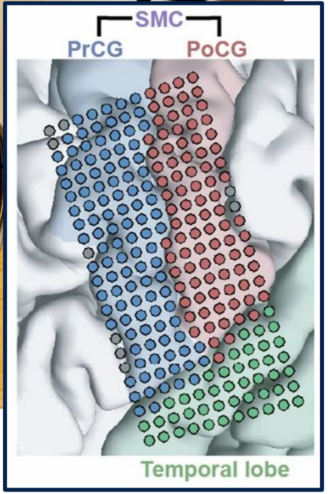
Metzger, Liu, Moses, et al. (2022)

High-performance BCI system for decoding extensive and multimodal speech

- + 1000 vocabulary
- Direct synthesis to voice
- Embodied decoding through avatar



Metzger*, Littlejohn*, Silva*, Moses*, Seaton*, et al., A high-performance neuroprosthesis for speech decoding and avatar control. Nature 2023



Participant: Ann

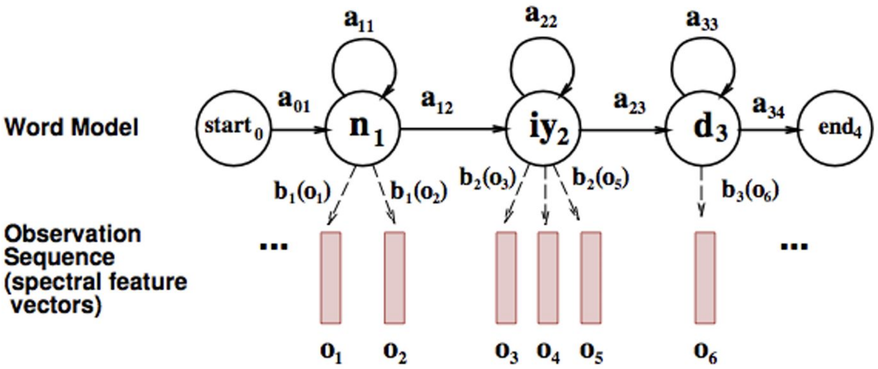
- Severe paralysis by brainstem stroke more than 20 years ago.
- Cannot speak or vocalize sounds.
- Cannot control orofacial movements.
- Attempting to mime a sentence prompted on the screen

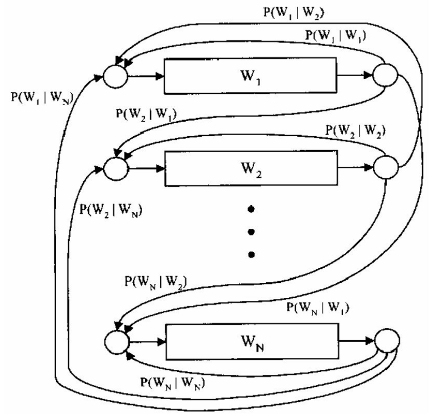
Challenge: ground truth behaviors are often not observable in participants with severe paralysis

- *Especially, **timing information (“when”)** is not accessible for participants who have minimal or no residual speech ability to vocalize.*

Similar setting as automatic speech recognition (ASR)

- ASR models (e.g., HMM): inferring temporally unaligned words from acoustic features
- Brain decoding: inferring temporally unaligned words from ***neural (brain)*** features





Training Counts

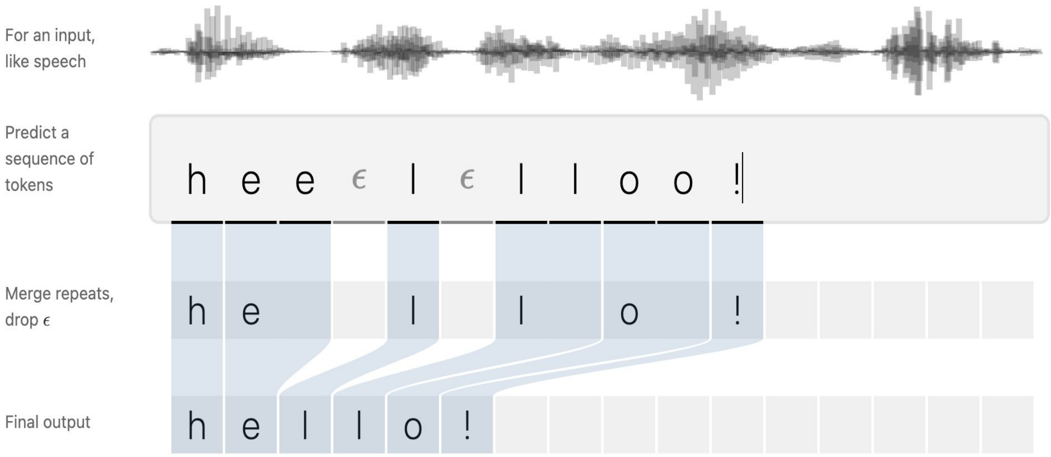
198015222 the first
194623024 the same
168504105 the following
158562063 the world
…
14112454 the door

23135851162 the *

$$\hat{P}(\text{door}|\text{the}) = \frac{14112454}{23135851162}$$

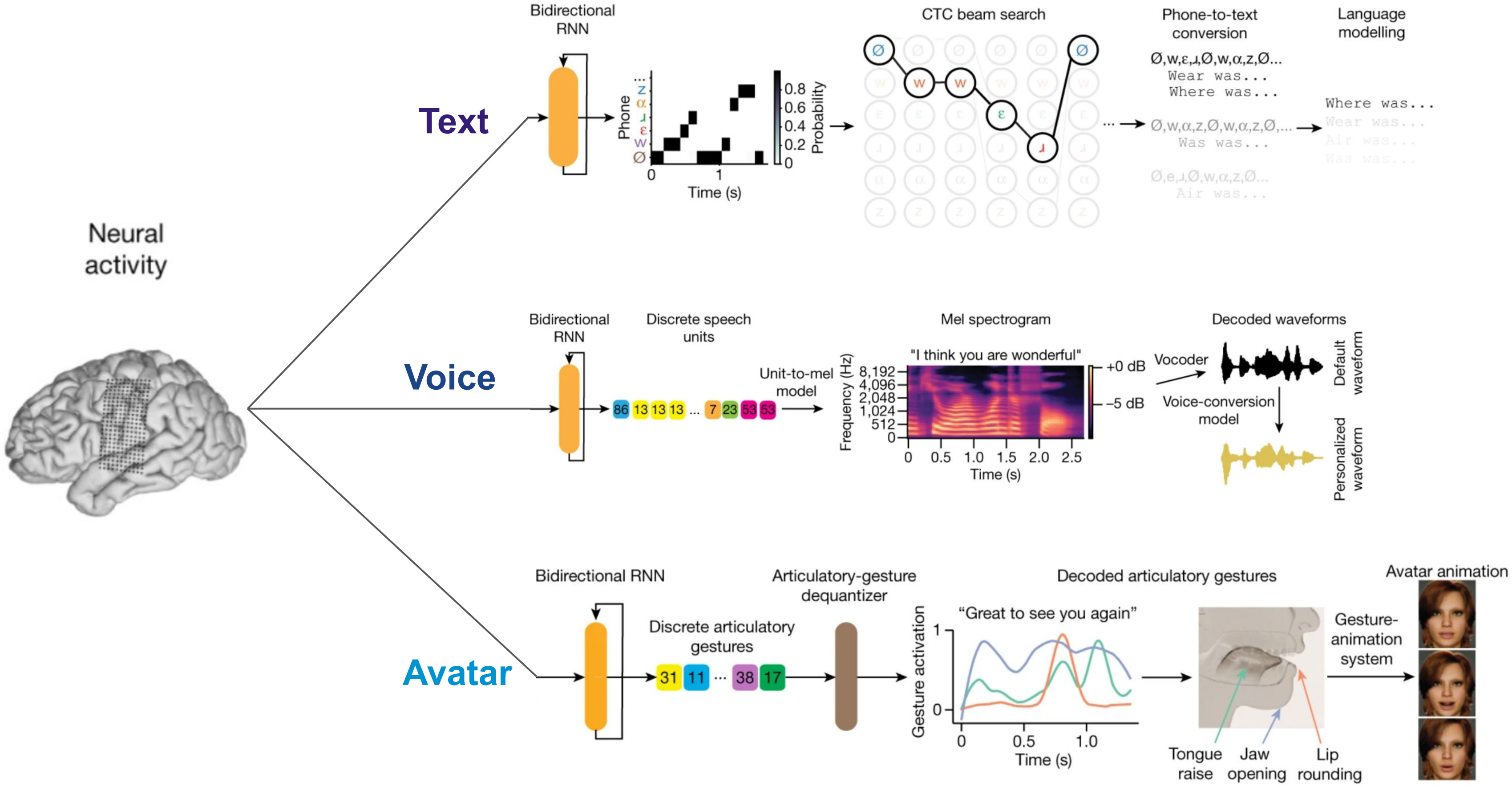
$$= 0.0006$$

Connectionist temporal classification (CTC) for training without time-aligned targets



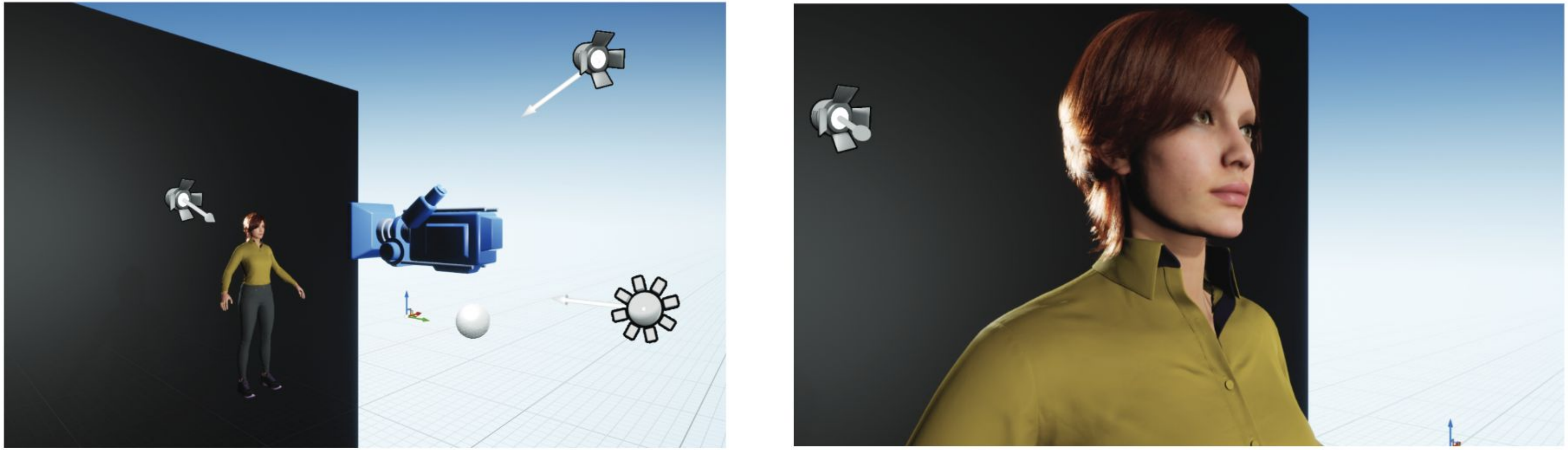
- CTC learns to dynamically infer alignment between frame-wise probabilities to unaligned target sequence.
- Successfully applied to several brain decoding studies (Sun et al., 2020), including clinical applications (Metzger et al., 2023; Willet et al., 2023).

Hannun, Sequence Modeling with CTC, Distill, 2017.



Metzger*, Littlejohn*, Silva*, Moses*, Seaton*, et al., A high-performance neuroprosthesis for speech decoding and avatar control. Nature 2023

Enabling virtual embodiment + interaction

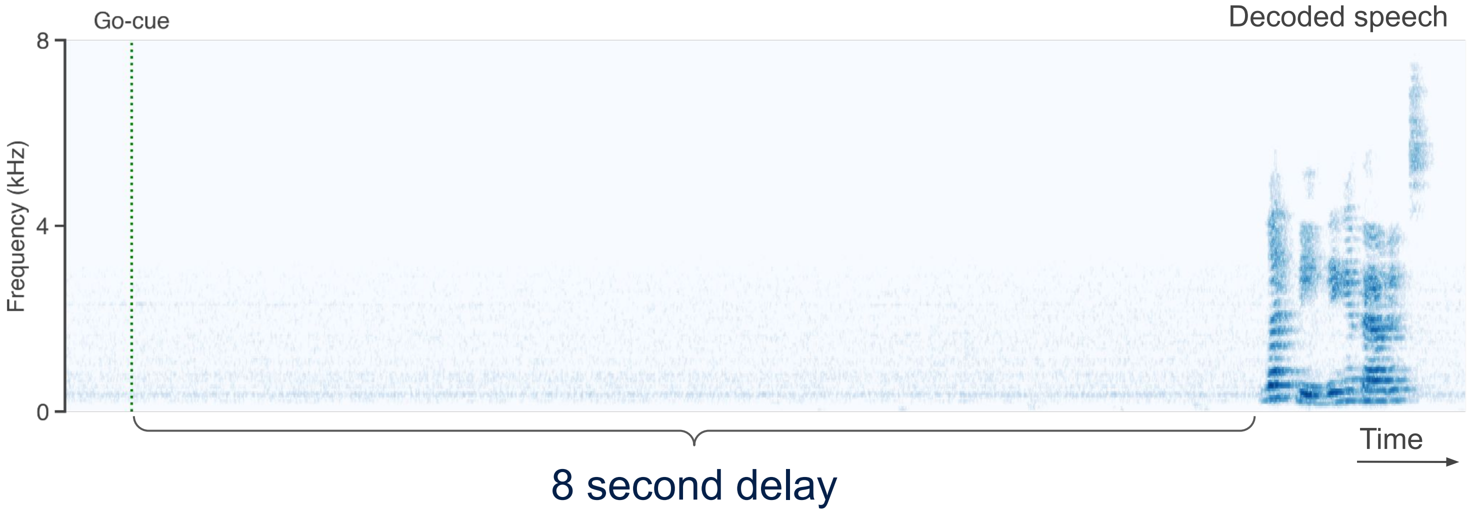


Virtual environment for avatar decoding

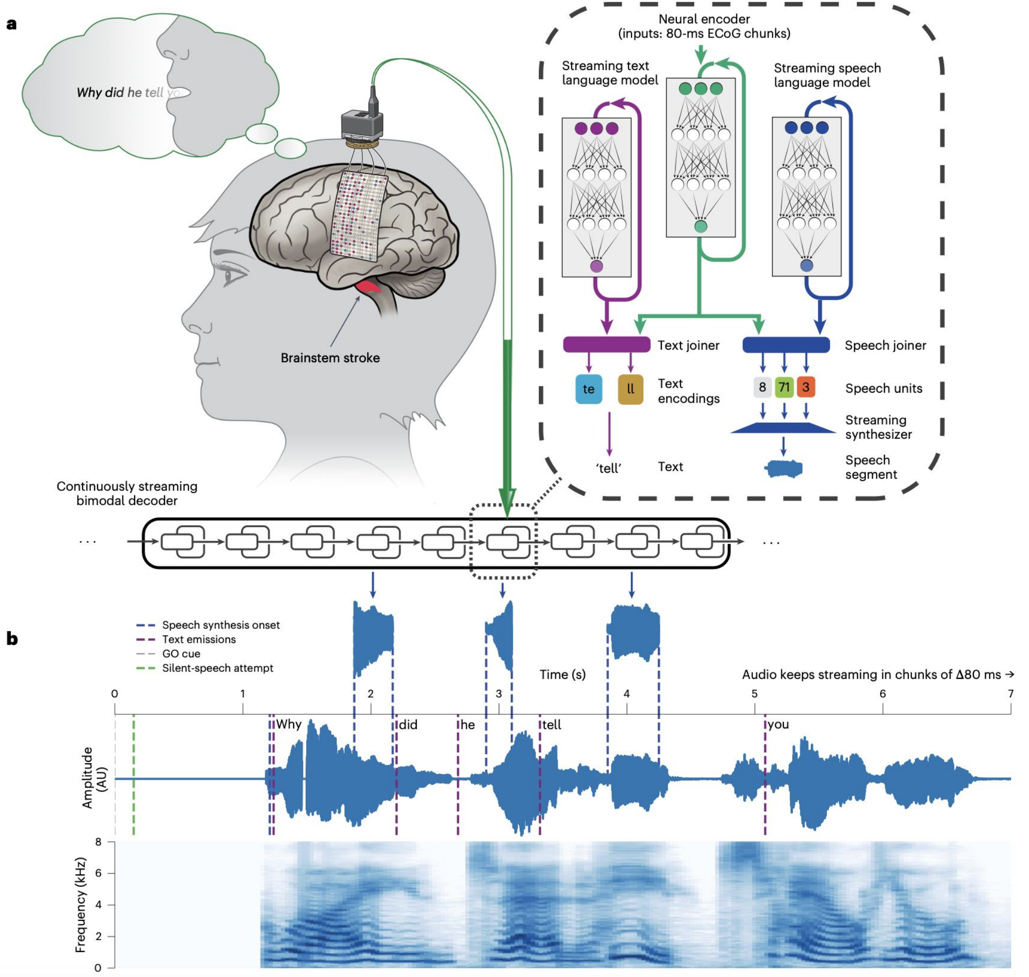
Sean L. Metzger*, **Kaylo T. Littlejohn***, Alexander B. Silva*, David A. Moses*, Margaret P. Seaton*, A high-performance neuroprosthesis for speech decoding and avatar control. Nature 2023

CTC decoding approach incurs a long delay

- Decoder synthesizes predicted speech based on an 8 second window, resulting in a long delay time



A streaming brain-to-voice neuroprosthesis to restore naturalistic communication

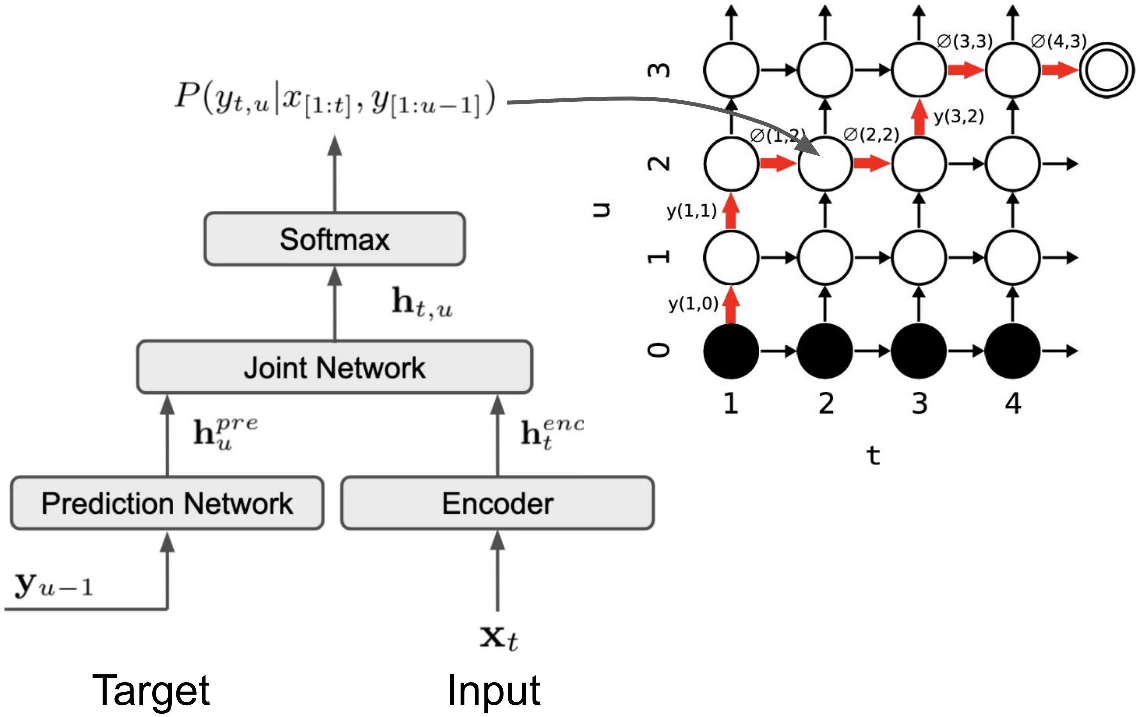


Littlejohn*, Cho*, …, Chang+, and Anumanchipalli+,
Nature Neuroscience, 2025

Limitations of the previous CTC for streaming decoding

- CTC lacks an architectural mechanism for streaming inference.
- CTC assumes conditional independence between output tokens.
  - No explicit language prior in training, which is crucial for accurate inference.
  - Thus, it typically requires a large window of inputs.

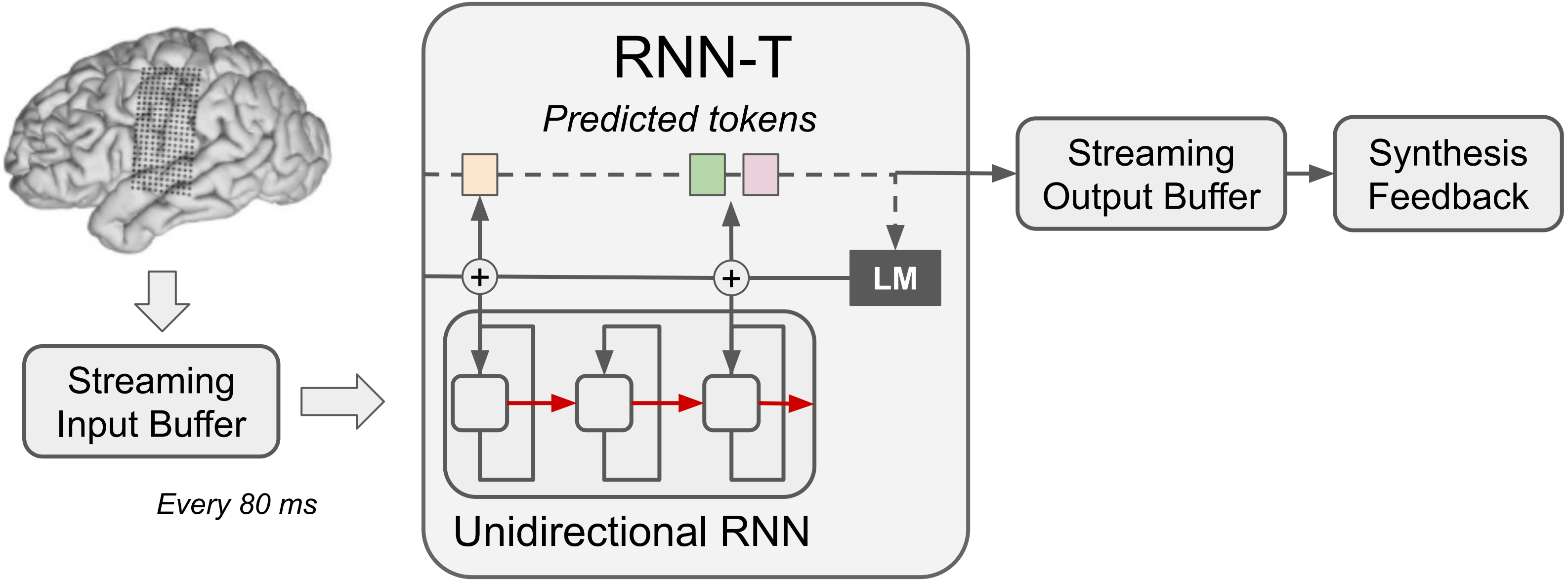
Recurrent neural network-Transducer (RNN-T)



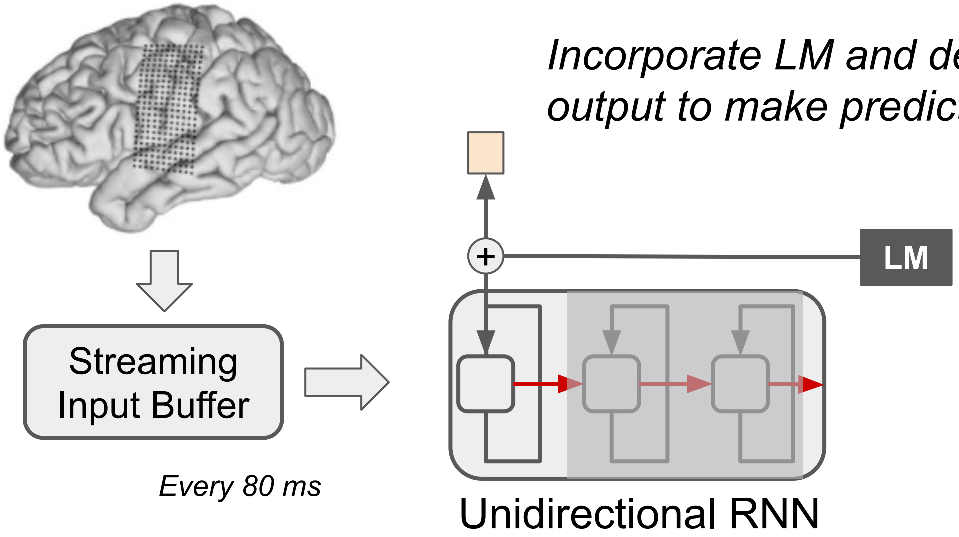
- ***A joint language model (predictor) imposes prior on the incremental prediction.***
- A dynamic path search on the alignment grid during training and inference.

Graves. Sequence Transduction with Recurrent Neural Networks. Arxiv 2012

A streamable speech BCI framework via RNN-T



Incorporate LM and decoder output to make prediction.



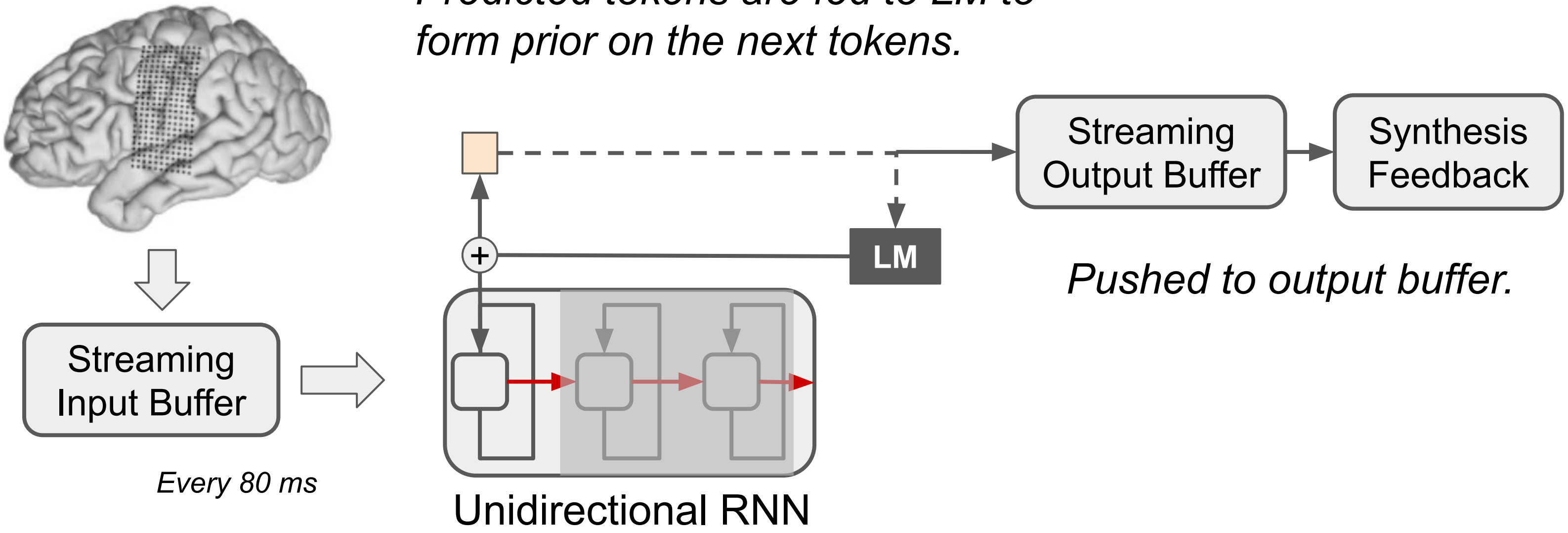
Streaming Input Buffer

Every 80 ms

LM

Unidirectional RNN

Predicted tokens are fed to LM to form prior on the next tokens.



Streaming Input Buffer

Every 80 ms

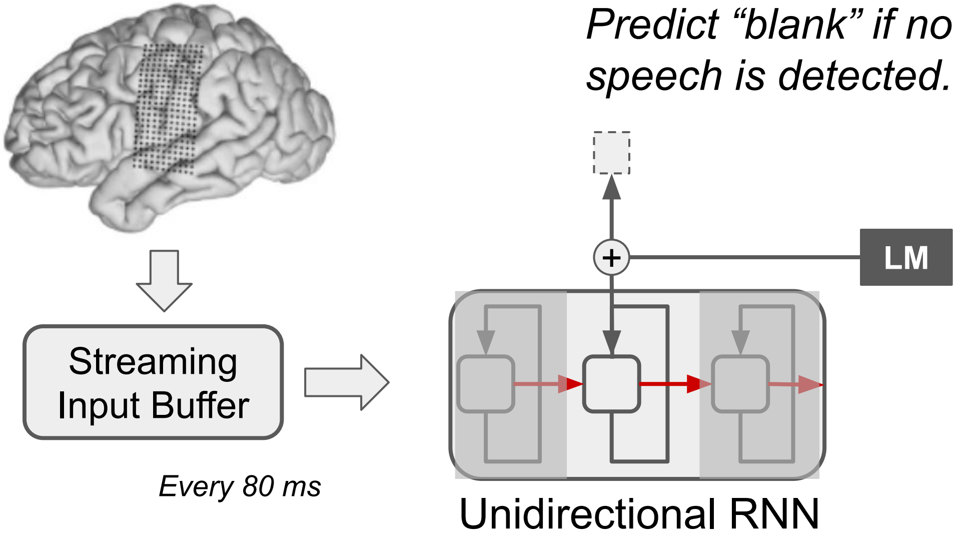
Unidirectional RNN

LM

Streaming Output Buffer

Synthesis Feedback

Pushed to output buffer.

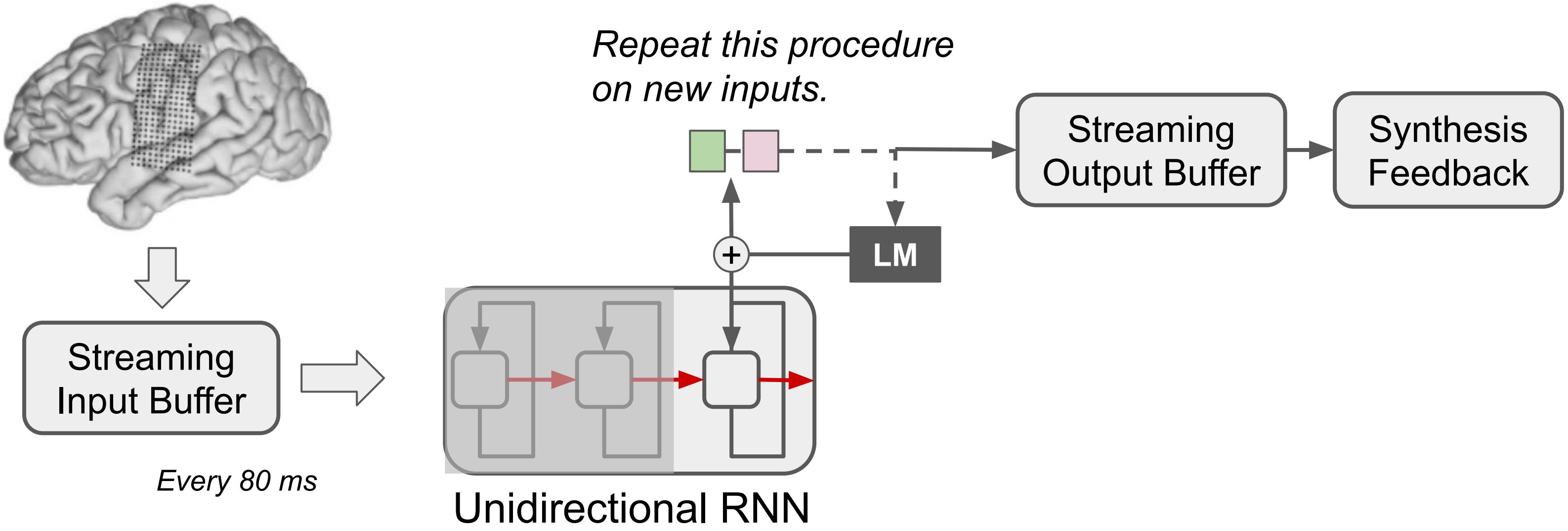


Predict “blank” if no speech is detected.

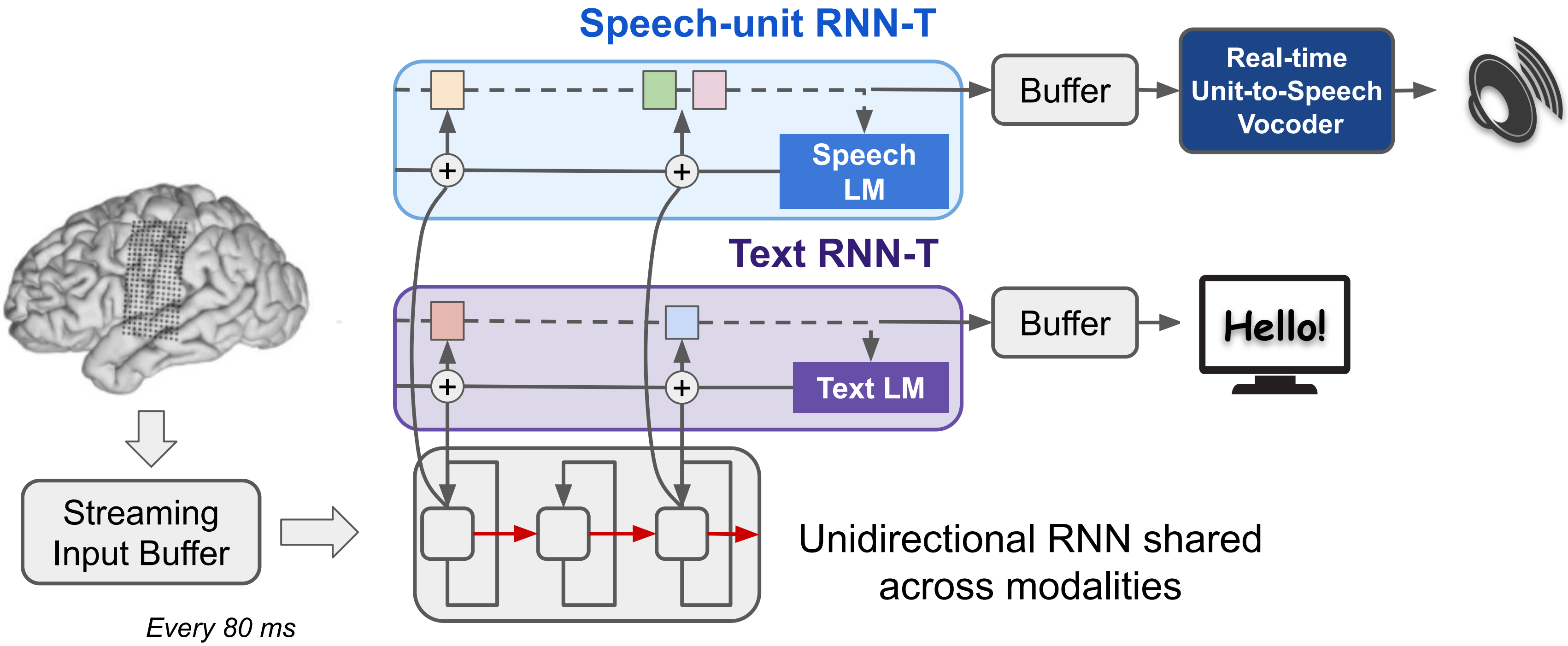
Streaming Input Buffer

Every 80 ms

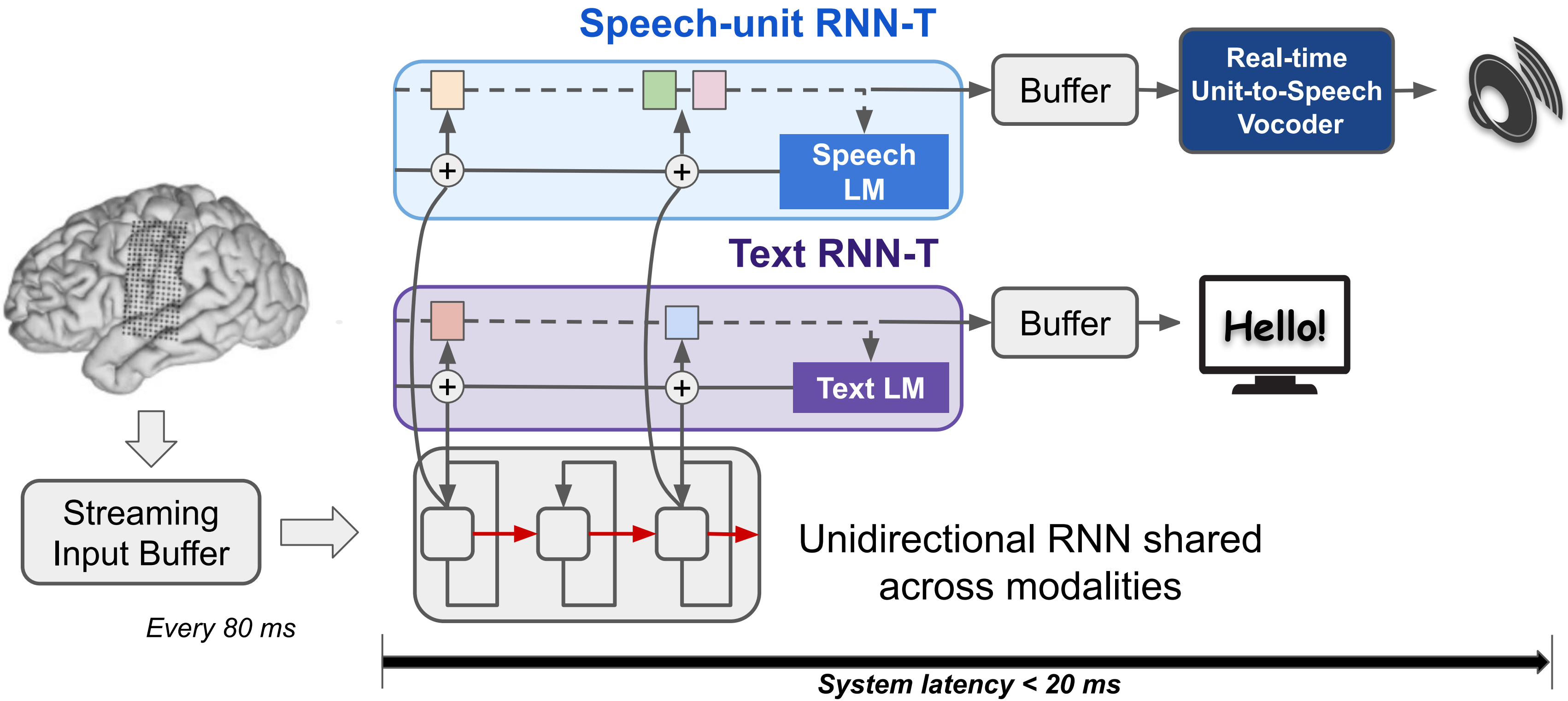
Unidirectional RNN


Repeat this procedure on new inputs.
Streaming Output Buffer
Synthesis Feedback
+
LM
Streaming Input Buffer
Every 80 ms
Unidirectional RNN

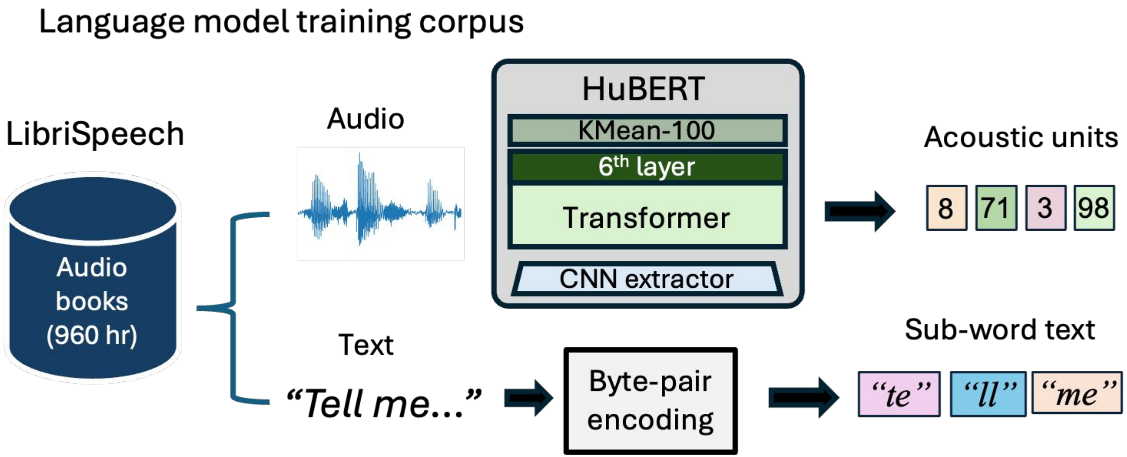
Multimodal RNN-T neural decoder: speech & text joint decoding



Multimodal RNN-T neural decoder: speech & text joint decoding



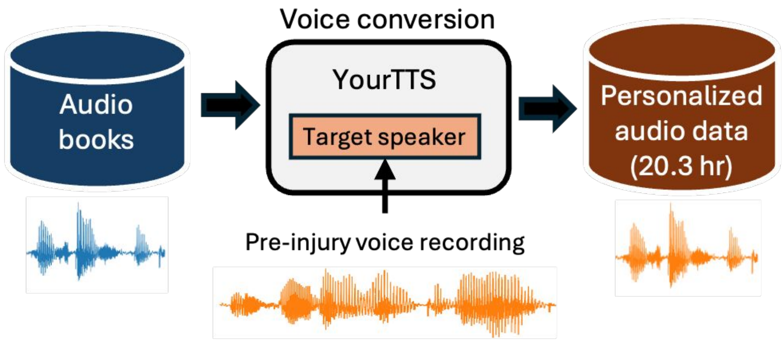
Pretraining LM on a large speech corpus to overcome the limited coverage of speech



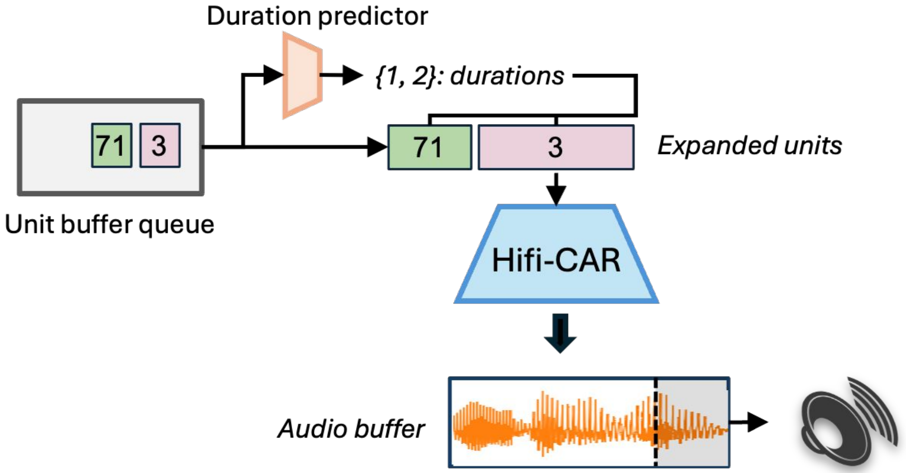
- RNN-T LMs were pretrained using a large speech audio dataset.

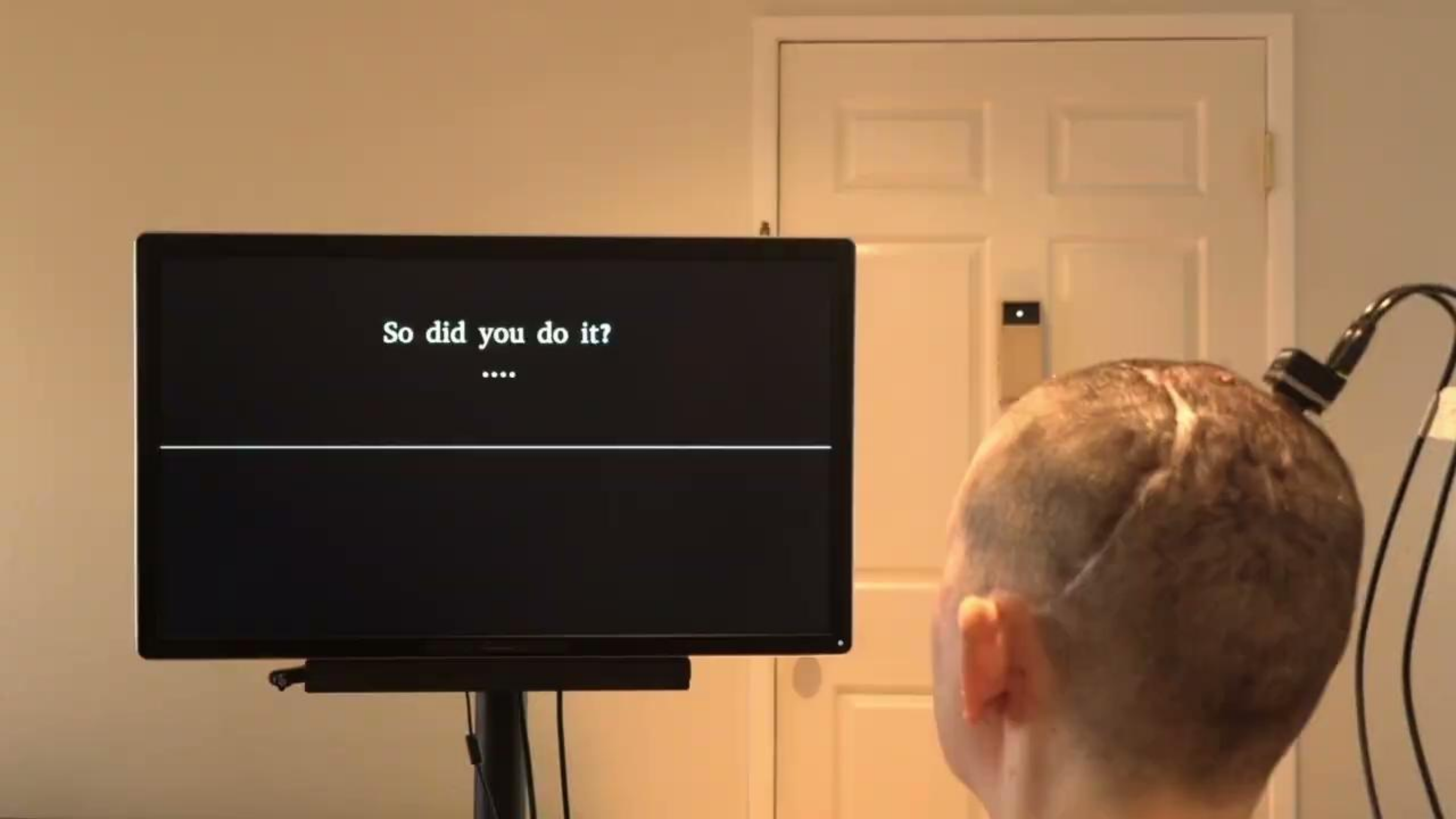
A personalized streamable synthesizer that restores participant’s original voice

Synthesizer training corpus



Streaming synthesizer

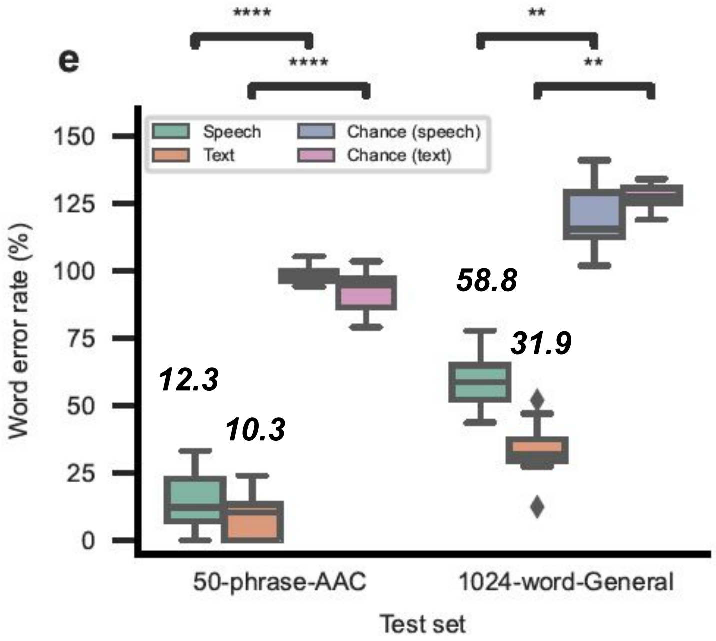


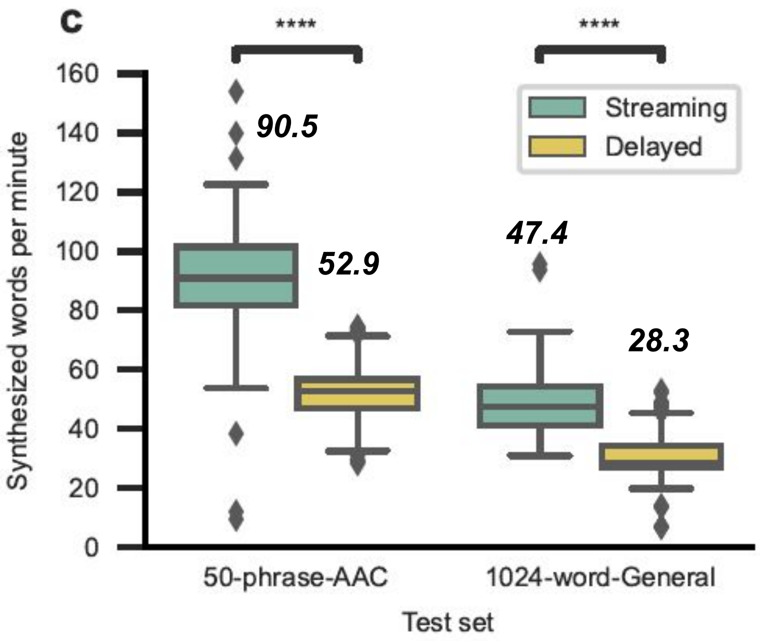


So did you do it?

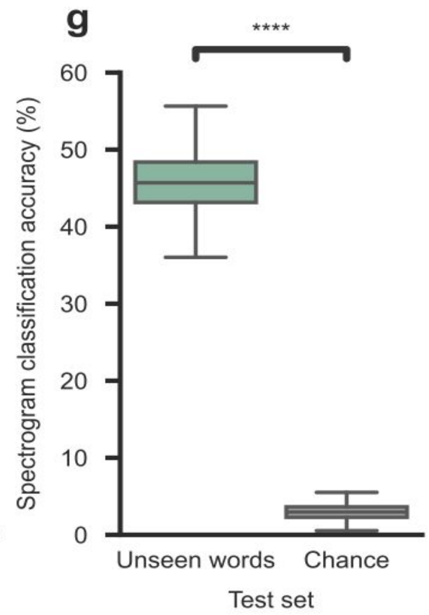
....

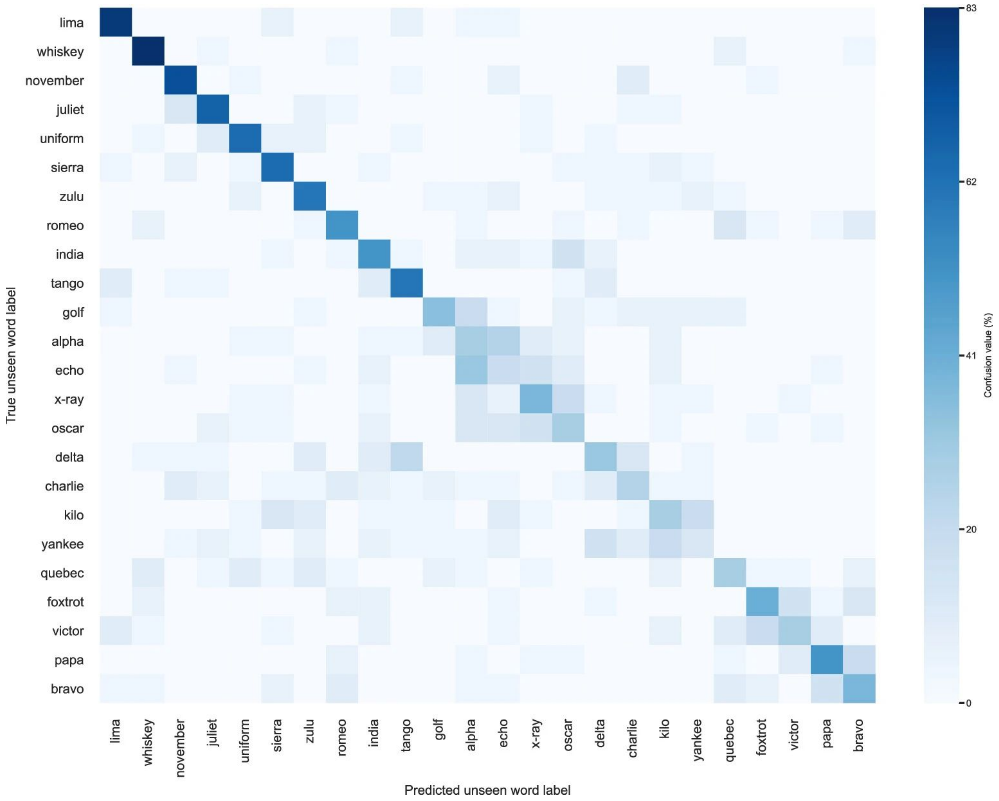
Fast streaming intelligible speech synthesis and text decoding





Generalization to rare unseens words





Future of communication neuroprostheses

1) Wireless, low footprint high-density devices

2) Closed-loop speech/avatar decoding

3) Universal vocal tract models grounded in speech science

4) Complete virtual embodiment (body / hand / voice)

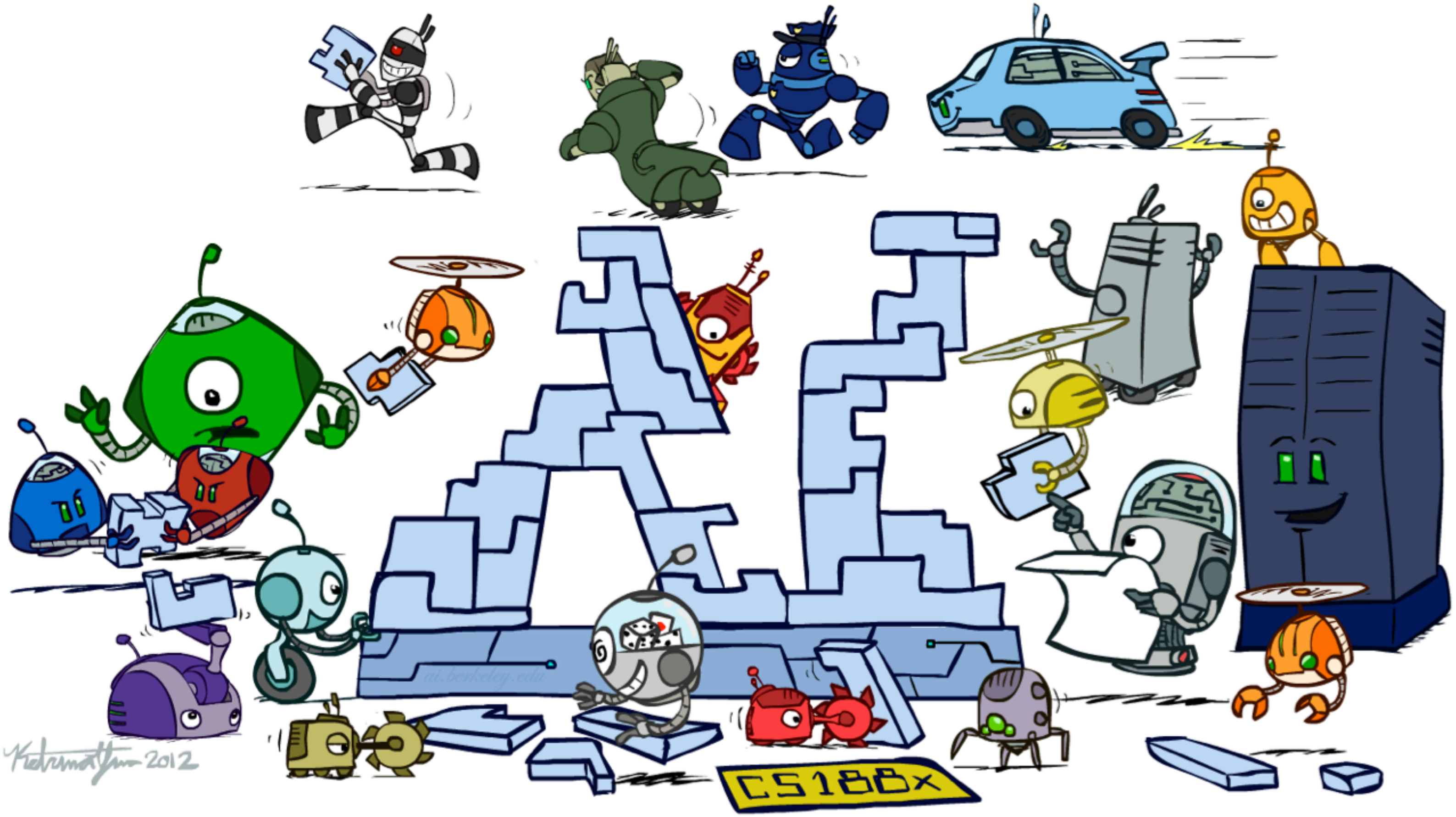
The Ideal Speech BCI

- Performance
  - 80 words per min
  - <5% word error rate
  - Easy to use by patient, caregivers, and communication partners
- Robustness
  - Easy and fast calibration
  - Long-term stability
  - Can be used across multiple indications (e.g., ALS, stroke, and others)
  - Performance retained in different settings (indoor and outdoor, daytime and nighttime) and over disease progression
  - Safeguards to maintain privacy of personal data
- Safety
  - Fully implantable, wireless
  - Safe and easy to implant and explant

Adapted from Edward Chang: Brain–Computer Interfaces for Restoring Communication. NEJM 2024.

CS 188: Artificial Intelligence

Fuzzy Logic



Instructor: Oliver Grillmeyer --- University of California, Berkeley

Announcements

- HW10 is due **Thursday, August 7**, 11:59 PM PT
- Project 5 is due **Friday, August 8**, 11:59 PM PT
- Ignore assessment on HWs part B, but please show your work
- Final Exam is **Wednesday, August 13**, 7-10 PM PT in 2050 VLSB
- Course evaluations open now

Fuzzy Logic

- History
- Fuzzy Sets
- Fuzzy Expert Systems

History

- Influenced by work in Multivalued Logic - Jan Lukasiewicz, 1920s
- Lotfi Zadeh (U.C. Berkeley professor) developed math for Fuzzy Sets, 1965
- Fuzzy Controller for steam engine, mid 1970s
- Fuzzy Expert System for mixing and grinding cement in cement kiln, 1982
- Fuzzy Controlled subway system in Japan, 1986
- Fuzzy commercial and industrial systems counts:
  - 1986: 8
  - 1991: 300
  - 1993: 1500

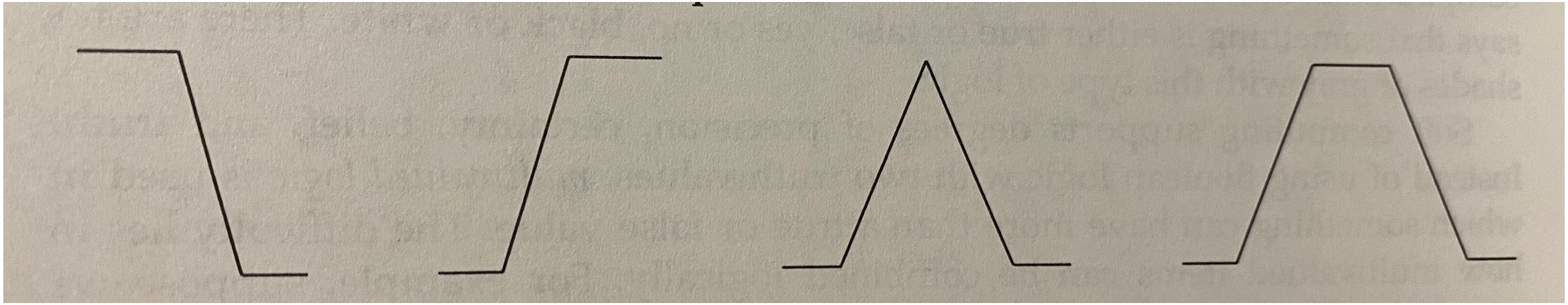
Fuzzy Sets

- Represent noncrisp values like Tall, Fast, Comfortable
- These have degrees of fit represented as Membership Grades
- Tall Fuzzy Set

Actual Height	Membership Grade
<= 5’3”	0
5’6”	0.25
5’9”	0.5
6’	0.75
>=6’3”	1

Fuzzy Set Functions

- Fuzzy Sets represented as functions
- Named after their shape: L, Γ, Λ, Π (‘L’, Gamma, Lambda, Pi)



Tall Fuzzy Set

- We can represent membership grade (degrees of fit) with functions
- Tall Fuzzy Set

Actual Height	Membership Grade
h <= 5’3”	0
5’3” <= *h* <= 6’3”	(*h* - 5’3”) / 12”
h >= 6’3”	1

Linguistic Variables and Hedges

- Linguistic Variables represented as fuzzy sets, e.g., *height* has value *Tall*
- Hedges modify Linguistic Variables, e.g., *Very, Somewhat, More or Less*
- *Very* can be represented by squaring the Membership Grade of *Tall*

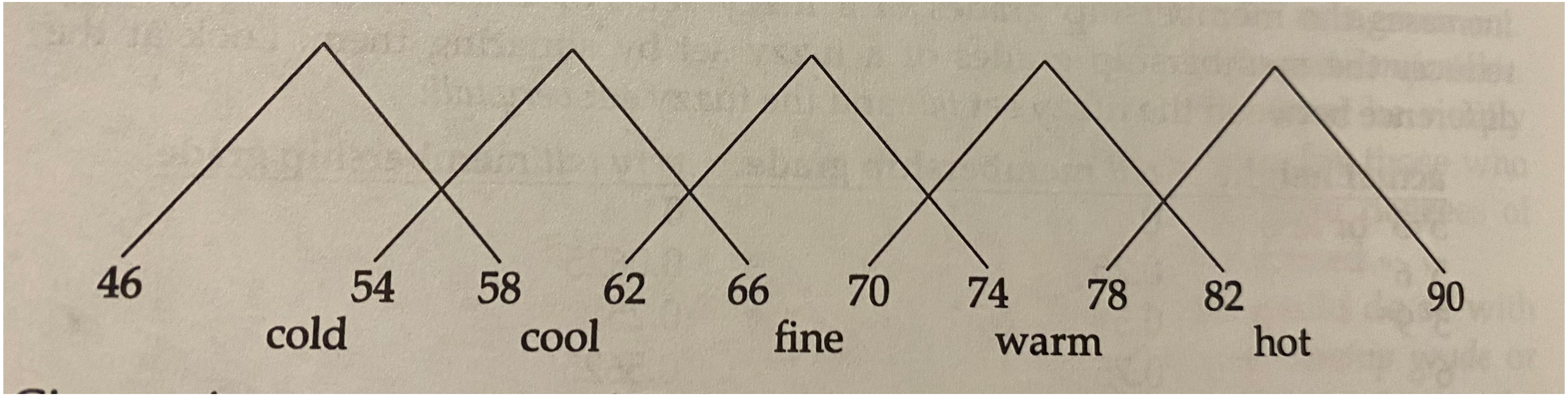
Actual Height	Tall Membership Grade	Very Tall Membership Grade
<= 5’3”	0	0
5’6”	0.25	0.0625
5’9”	0.5	0.25
6’	0.75	0.5625
>=6’3”	1	1

Fuzzy Sets for a Fan Controller

- Linguistic Variables and their values
  - *temperature*: *Cold, Cool, Fine, Warm, Hot*
  - *humidity*: *Low, Medium, High, Very High*
  - *fan-speed*: *Slow, Medium, Fast*

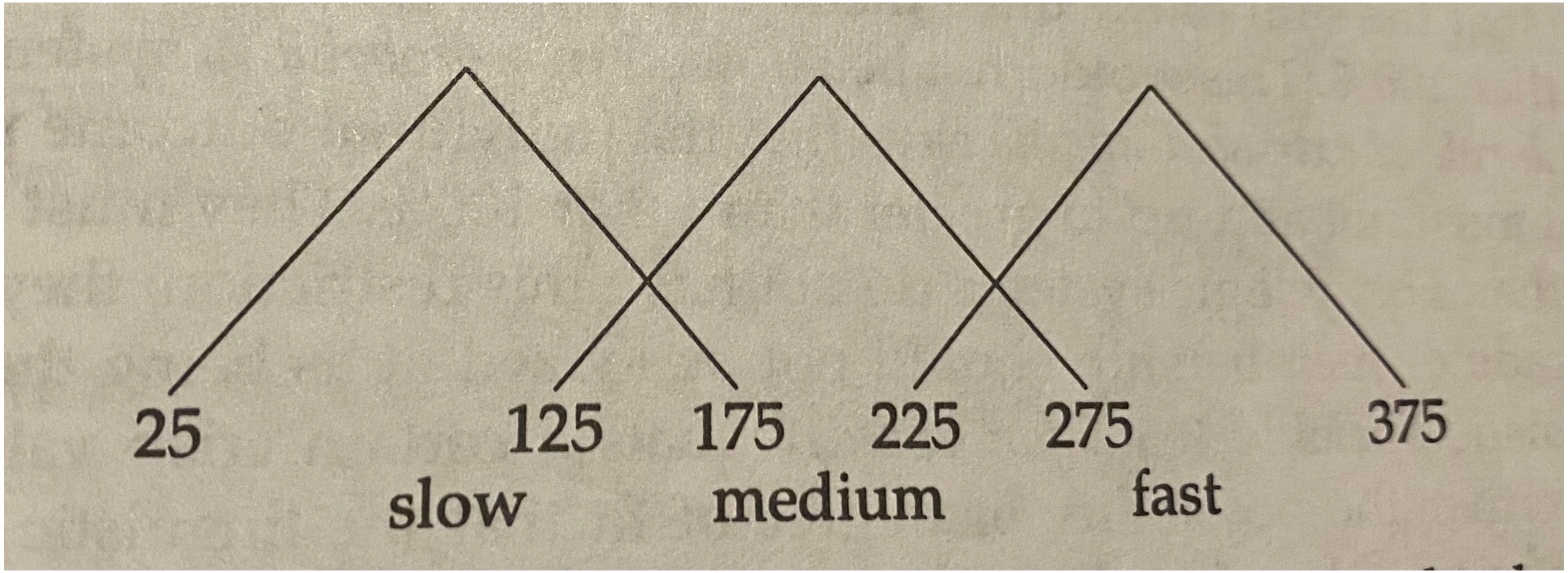
Temperature Fuzzy Set

- Linguistic Variables and their values
  - *temperature*: *Cold, Cool, Fine, Warm, Hot*
- Overlapping sets is important



Fan-Speed Fuzzy Set

- Linguistic Variables and their values
  - *fan-speed*: *Slow, Medium, Fast*

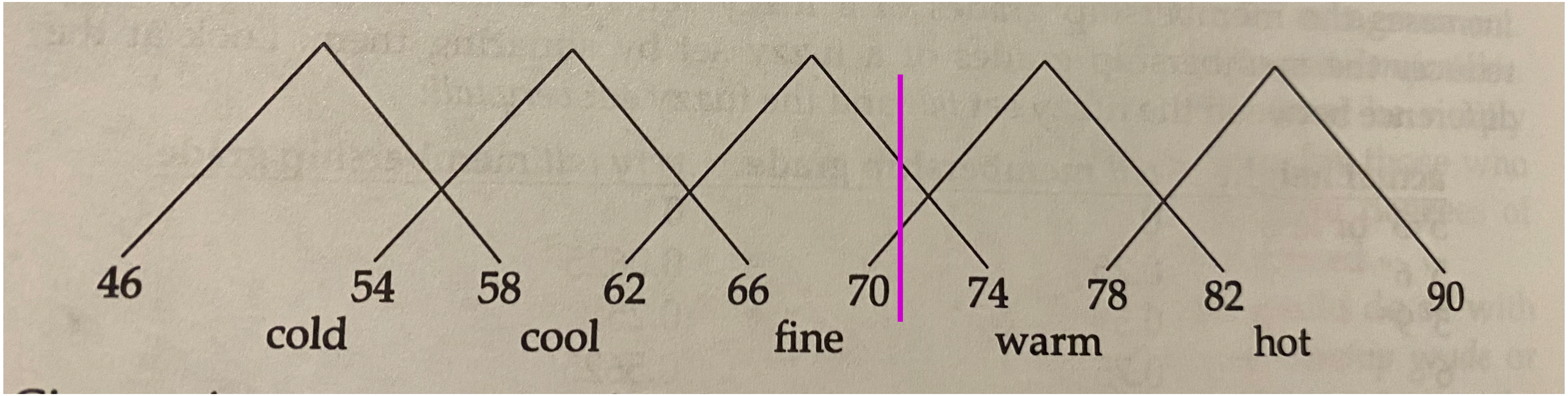


Fuzzy Controller Process

- Get crisp inputs (typically from sensors)
- Fuzzify crisp inputs
- Match fuzzy inputs to rules
- Combine rules according to membership grades to generate fuzzy output sets
- Defuzzify output sets to get crisp values
- Apply crisp values to the system being controlled

Fuzzify Temperature Crisp Input

- Assume: actual temperature is 71 degrees
- Match with *temperature* fuzzy sets: *Fine* and *Warm*
  - Membership grade in *Fine* is 0.5
  - Membership grade in *Warm* is 0.167



Fuzzify Humidity

- Assume: humidity is 65%
- Match with *humidity* fuzzy sets: *Medium* and *High*
  - Membership grade in *Medium* is 0.8
  - Membership grade in *High* is 0.3

Fuzzy Rules

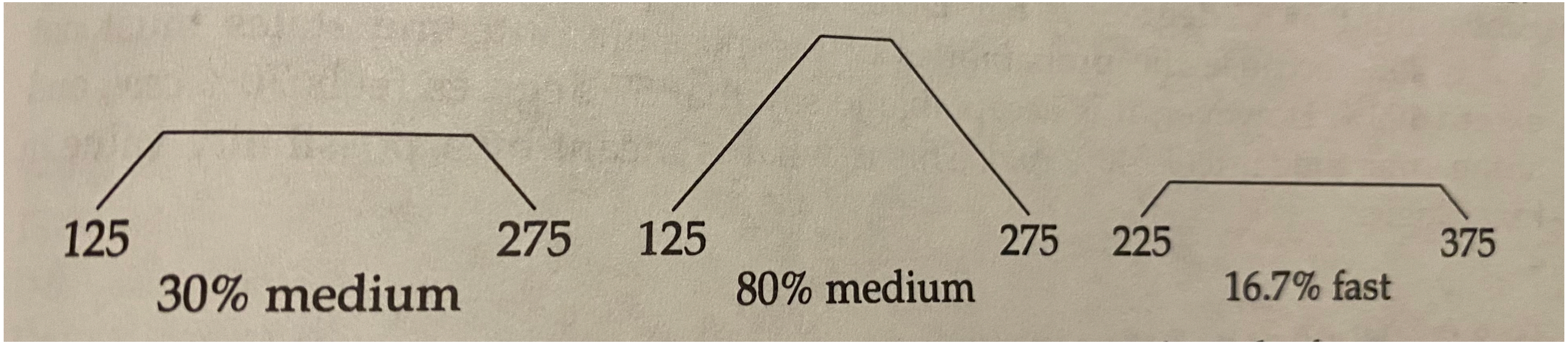
- Combining Rule-based system with Fuzzy Logic and Fuzzy Sets
- Can represent statements with a degree of imprecision
    - if *temperature* is *Cool*, then set *fan-speed* to *Slow*
    - if *temperature* is *Warm*, then set *fan-speed* to *Fast*
    - if *temperature* is *Fine* and *humidity* is *High*, then set *fan-speed* to *Medium*
    - if *temperature* is *Warm* or *humidity* is *Medium*, then set *fan-speed* to *Medium*

Match Fuzzy Inputs to Rule Conditions

- *temperature* is 0.5 *Fine* and 0.167 *Warm*
- *humidity* is 0.8 *Medium* and 0.3 *High*
- Match against condition of rules according to degree of fit
  - 0: if *temperature* is *Cool*
  - 0.167: if *temperature* is *Warm*
  - min(0.5, 0.3): if *temperature* is *Fine* and *humidity* is *High*
  - max(0.167, 0.8): if *temperature* is *Warm* or *humidity* is *Medium*

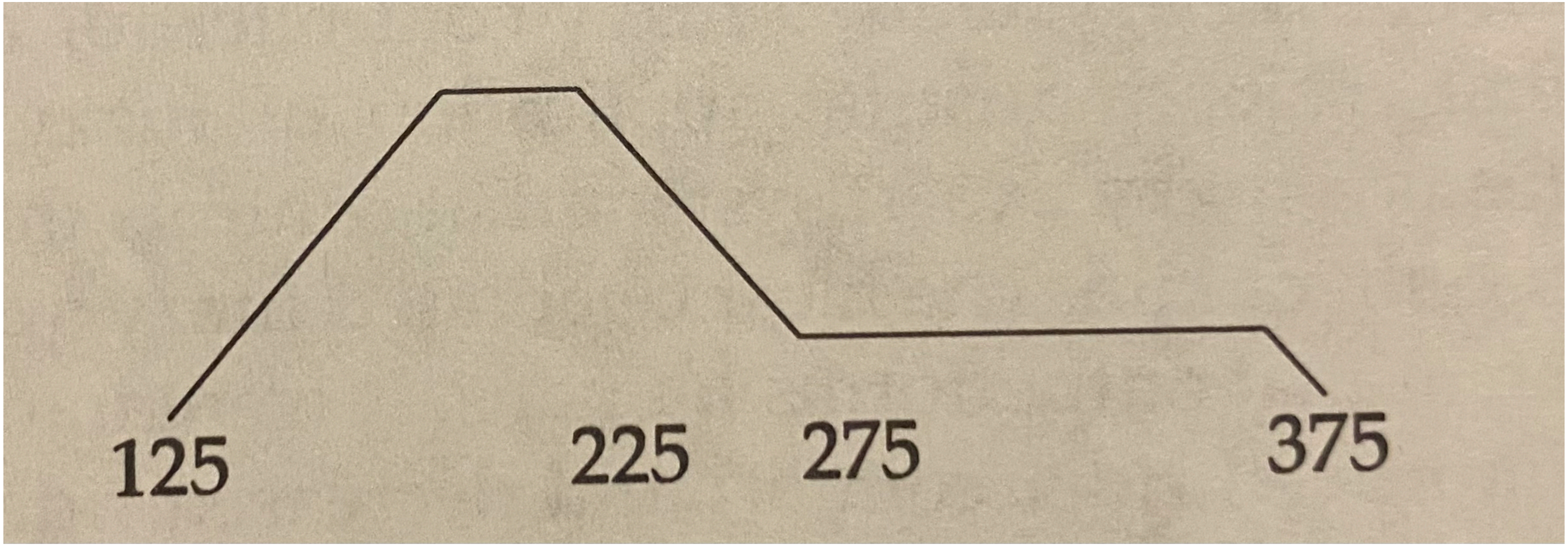
Apply Rule Actions

- Apply membership grade from condition to weigh effect of action
  - if *temperature* is *Warm* -> <u>set *fan-speed* to Fast (16.7%)</u>
  - if *temperature* is *Fine* and *humidity* is *High* -> <u>set *fan-speed* to *Medium* (30%)</u>
  - if *temperature* is *Warm* or *humidity* is *Medium* -> <u>set *fan-speed* to *Medium* (80%)</u>



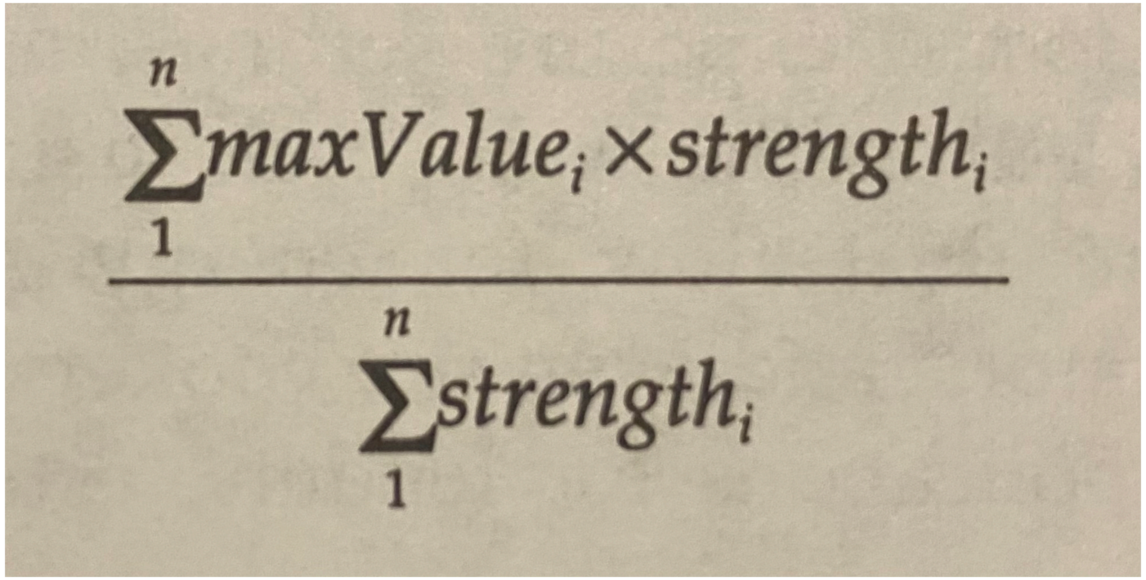
Combine Rule Actions

- Create new fuzzy sets representing weighted applicable actions
  - Sum weights or take maximum (maximum used here)



Defuzzify Output Fuzzy Set

- Produce a crisp value from the Fuzzy Set
- Different methods:
  - Center of mass or Center of gravity
  - Average of Maximums
    - max of *Medium*: 200 RPM
    - max of *Fast*: 300 RPM
- (200 x 0.8 + 300 x 0.167)/(0.8 + 0.167)
- 217.27 RPM



$$\frac{\sum_{1}^{n} maxValue_i \times strength_i}{\sum_{1}^{n} strength_i}$$

Possibility vs Probability

- Recall that our membership grades for 65% *humidity* were 0.8 *Medium* and 0.3 *High*
- Also our output sets were 80%, 30%, and 16.7%
- These exceed 100%
- Problem for probability
- Okay in Fuzzy Logic because they are Possibilities
- Don’t say probability of 65% humidity = medium is 80% and high is 30%
- Instead say 65% humidity feels like 80% medium humidity and 30% high
- Temperature example: 57 degrees feels 50% Cool and 17% Cold