Speech Brain-Computer Interfaces for Restoring Natural Communication

Past, now, and future

Chaofei Fan, CS224N
DOING ACTIVITIES HAVING LESS ON YOUR MIND
What we take for granted is lost for some individuals

• Howard Wicks, 21, lost all his dreams after a severe stroke.

• Neurological disorders like brainstem stroke or Amyotrophic Lateral Sclerosis (ALS) can cause speech and motor impairment and even complete loss of speech.

• These individuals are facing extreme challenges in their lives.

• Communication with loved ones and caretakers is one of their most desperate needs.
Assistive communication devices
Neuralink Prime Study
Tap into the intact mind with brain-computer interfaces (BCIs)

“[The Link] has helped me reconnect with the world, my friends, and my family. It's given me the ability to do things on my own again without needing my family at all hours of the day and night.”

— Noland Arbaugh, PRIME Study participant
A Brief History of BCI
How does BCI work? Start from the beginning...

Electricity in the brain
Listen to the brain from the outside
Electroencephalogram (EEG)

Hans Berger, a German psychiatrist, invented EEG and succeeded in recording the first human EEG in 1924.
Perform with brain waves

Electroencephalogram (EEG)
Listen to single neurons in motor cortex

Hubel 1988
Neurons communicate with spikes

Goodman, Spiking Neural Networks
Single neurons encode movement directions

A Single neuron, multiple trials

Shenoy & Yu, Brain Machine Interfaces
Multiple neurons for accurate decoding

A One neuron

Firing rate (spikes/s)

0 60 120 180 240 300 360
Direction (degrees)

Neuron 1 activity

Shenoy & Yu, Brain Machine Interfaces
Multiple neurons for accurate decoding

A Calibration phase

Boundaries moved to optimize discrimination

Shenoy & Yu, Brain Machine Interfaces
Techniques for measuring neural activity


Harvard/MGH-NMR.
Beyond 2D control

Control robotic arms
Beyond 2D control

Restore walking
Beyond 2D control
Communication through handwriting

Willett et al. 2021
90 char/min with 95% accuracy

Communication through handwriting
Restore natural speech?

- Sip and puff interface
- BCI-driven cursor control
- Handwriting
- QWERTY touch screen
- Touch screen and predictive text
- Eye tracking and predictive text
- Professional typewriting
- Presentation-style speech
- Conversational speech

Brain-to-text iBCI

2D cursor iBCI
Pandarinath*, Nuyujukian*, ..., Hochberg, Shenoy**, Henderson** (2017) eLife

Chang & Anumanchipalli (2019) JAMA
Language processing in the brain

**Perception**
Perception of the surface properties of linguistic input (for instance, speech perception area)

**Motor planning**
Planning of the motor movements needed to realize linguistic output (for instance, Broca’s area)

**Language**
Language knowledge and processing (language network)

**Knowledge and reasoning**
- Task demands beyond language (multiple demand network)
- Pragmatics, social reasoning (theory of mind network)
- Narratives, situation modelling (default mode network)

**Intended meaning**
(multiple brain areas, including the above)

Fedorenko et al. 2024
How do we produce speech sound
Motor cortex encodes articulatory and phonemic information
Small vocabulary speech BCI with ECoG

Moses et al. 2021
A High-Performance Speech Neuroprosthesis
Implanting microelectrode arrays into BrainGate2 clinical trial participant T12

- T12 has bulbar-onset Amyotrophic Lateral Sclerosis (ALS)
- She retains some limited orofacial movement and an ability to vocalize, but is unable to produce intelligible speech.
- Four 64-channel Utah arrays
  - Two in area 6v (ventral motor cortex)
  - Two in area 44 (part of Broca’s area)

Willett, Kunz, Fan, et al. 2023
Neural representation of orofacial movements and speech

Willett, Kunz, Fan, et al. 2023
Real-time brain-to-text BCI

Willett, Kunz, Fan, et al. 2023
I don't want to call her a baby sitter.
I do not have much to compare it to.
Data collection
Data collection

Training and evaluation sentences are randomly selected from the Switchboard corpus of telephone conversation (~1,0000 sentences)
Problem definition

- Neural feature inputs: \( \{x_1, x_2, \ldots, x_n\} \quad x_i \in \mathbb{R}^{d \times 1} \)
- Words outputs: \( \{y_1, y_2, \ldots, y_m\} \quad y_i \in \mathbb{R}^{v \times 1} \)
Problem definition

- Neural feature inputs: \( \{x_1, x_2, \ldots, x_n\} \quad x_i \in \mathbb{R}^{d \times 1} \)

- Words outputs: \( \{y_1, y_2, \ldots, y_m\} \quad y_i \in \mathbb{R}^{V \times 1} \)

Diagram:

- \( x_1 \quad x_2 \quad \ldots \quad x_n \)
- \( \alpha \quad \text{SIL} \quad k \quad \ae \quad n \quad \text{SIL} \quad s \quad p \quad i \quad k \quad \text{SIL} \)
- Phonemes to words decoder
- Neural to phonemes decoder

\( l \quad \text{can} \quad \text{speak} \)
Neural to phonemes decoder

• A Seq2Seq problem

• Encoder-decoder models allow arbitrary alignment between inputs and outputs.

• Neural to phonemes decoding only needs monotonic alignment.
Sequence modeling with Connectionist Temporal Classification (CTC)

Handwriting recognition: The input can be \((x, y)\) coordinates of a pen stroke or pixels in an image.

Speech recognition: The input can be a spectrogram or some other frequency based feature extractor.
Sequence modeling with CTC

Inputs

CTC outputs post-processing

h h e ℓ l l l ℓ l l o
CTC training

The CTC conditional probability marginalizes over the set of valid alignments computing the probability for a single alignment step-by-step.

\[ p(Y \mid X) = \sum_{A \in \mathcal{A}_{X,Y}} \prod_{t=1}^{T} p_t(a_t \mid X) \]
What neural network to use?

Neural to phonemes decoder

$\alpha1$ SIL $k$ $\text{æ}$ $n$ SIL $s$ $p$ $i$ $k$ SIL

$x_1$ $x_2$ $\ldots$ $x_n$
What neural network to use?

Transformer

- 👍 Large datasets
- 👏 Long-range dependency

RNN

- 👉 Small datasets
- 👉 Short-range dependency
- 👉 Efficient real-time processing
Long Short-Term Memory (LSTM)

You can think of the LSTM equations visually like this:

\[ y = \text{softmax}(U h + b_2) \in \mathbb{R}^{|V|} \]

The + sign is the secret!

Source: http://colah.github.io/posts/2015-08-Understanding-LSTMs/
Gated Recurrent Units (GRU)
CTC inference

Y* = arg max P(Y | X)

Beam Search

Phoneme probability

CTC inference

GRU

X₁ X₂ ... Xₙ
CTC beam search

A standard beam search algorithm with an alphabet of \( \{\epsilon, a, b\} \) and a beam size of three.
The CTC beam search algorithm with an output alphabet \( \{\epsilon, a, b\} \) and a beam size of three.
CTC inference with language models

I can speak

Beam Search

Word insertion bonus

\[
Y^* = \arg\max_Y P(Y | X)
\]

\[
\approx \arg\max_Y P(Y | X)^\alpha \times P(Y) \times L(Y)^\gamma
\]

Probability of a sentence

\[
P(Y) = P(y_1)P(y_2 | y_1)P(y_3 | y_2, y_1) \ldots
\]

Phoneme probability

<table>
<thead>
<tr>
<th>SIL</th>
<th>SIL</th>
<th>SIL</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>k</td>
<td>k</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>n</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>s</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aɪ</td>
<td>aɪ</td>
<td>aɪ</td>
</tr>
<tr>
<td>æ</td>
<td>æ</td>
<td>æ</td>
</tr>
<tr>
<td>ε</td>
<td>ε</td>
<td>ε</td>
</tr>
</tbody>
</table>
Integrating language models in real-time decoding

\[ x_t \]

20ms time bin

Keep top-k hypotheses for next time bin

\[
\begin{align*}
P(I) &= 0.9 \\
P(EYE) &= 0.01
\end{align*}
\]
Transformer LM for 2nd pass rescoring

I can speak
I can spoke
...

Transformer LM

P(I can speak) = 0.95
P(I can spoke) = 0.01
...

I can speak

n-best hypotheses
Putting everything together
**Evaluation**

**Word error rate:** normalized edit distance between predicted words and ground truth words.

\[
WER(Y, \hat{Y}) = \frac{\text{distance}(Y, \hat{Y})}{\text{length}(Y)}
\]

Brain-to-Text Benchmark '24
What T12 said

“So many years of not being able to communicate and then suddenly the people in the room got what I said. I don’t remember what I exactly said after the prescribed script finished, but it had to be along the lines of ‘Holy shit, it worked, I’m so happy, and you guys did it.’”
Future of Speech BCIs
Multimodal Speech BCI
An accurate speech BCI for personal use

Card et al. 2024
Restoring effortless and natural communication by decoding inner speech

Decoding Accuracy (%)

- Attempted
- Miner
- Motoric Inner Voice
- Auditory Inner Voice
- Imagined Listening
- Listening
- Silent Reading

Decoding Accuracy (t12)

- 6v ventral
- 6v dorsal

Cope & Kalantzis / Transpositional Grammar • meaninacoutta.net/inner-speech

Erin, unpublished
Language processing in the brain

**Perception**
Perception of the surface properties of linguistic input (for instance, speech perception area)

**Motor planning**
Planning of the motor movements needed to realize linguistic output (for instance, Broca's area)

**Language**
Language knowledge and processing (language network)

**Knowledge and reasoning**
- Task demands beyond language (multiple demand network)
- Pragmatics, social reasoning (theory of mind network)
- Narratives, situation modelling (default mode network)

**Intended meaning**
Multiple brain areas, including the above

Fedorenko et al. 2024
BCIs raise new neuroethics considerations

• Should BCIs be allowed to read inner thoughts and memories?
• Read out inner speech that would not naturally be enacted?
• Read out memories that may otherwise be lost to Alzheimer’s disease?
• Read out subconscious fears to assist desensitization psychotherapy?
BCIs raise new neuroethics considerations

• Should BCIs be allowed to enhance cognitive function beyond natural levels?
  • Move a robotic arm faster and more accurately than a native arm?
  • Purchase a memory to skip a grade of mathematics in high school?
• We are currently grappling with the same questions:
  • Steroids, stimulants, elective plastic surgery ...
BCIs raise new neuroethics considerations

Although some of these ideas and questions may appear farfetched at present, as brain function and dysfunction continues to be revealed, BCI systems could build on these discoveries and create even more daunting ethical quandaries. But equally important is the immediate need to help people suffering from profound neurological disease and injury through restorative BMIs. In order to achieve the right balance it is imperative that we as physicians, scientists and engineers proceed in close conversation and partnership with ethicists, government oversight agencies, and patient advocacy groups.
Summary

• Recent advancements in AI and NLP, combined with years of neuroscience and neuroengineering research, show potential for restoring natural communication to people with speech impairments.

• We will soon have systems to assist people with communication disorders and paralysis.

• And understand better how the brain processes language!

• This brings hope to people like Howard and T12!
NEW HOPE FOR ALS PATIENTS
Acknowledgements

Krishna Shenoy
(9/3/1968 - 1/21/2023)

Howard Hughes Medical Institute

Participants
T5, T6 and T12

Stanford Neural Prosthetics Translational Lab (NPTL)

Funding
ALS Association
Stanford: Bio-X, Wu Tsai Neuroscience, OPA
NIH: NINDS, NIDCD, NICHD
VA Rehab. R&D Service
Mass. General Hospital ECOR
Larry and Pamela Garlick
Samuel and Betsy Reeves
John Gunn