Speech synthesis from neural decoding of spoken sentences

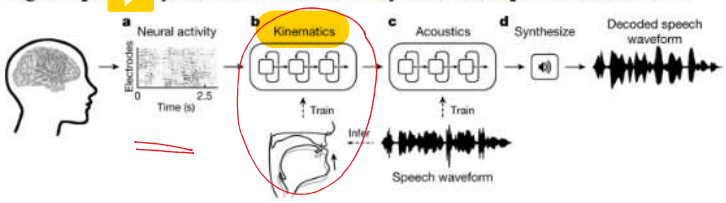
- Gopala K. Anumanchipalli,
- Josh Chartier &
- Edward F. Chang

Nature **volume 568**, pages493–498 (2019)

Nature.com, April 2019

From <https://www.nature.com/articles/s41586-019-1119-1>

Fig. 1: Speech synthesis from neurally decoded spoken sentences.



Neural activity in the cortex can be translated into speech using a synthetic device.

From <https://www.nature.com/articles/s41591-019-0488-1>

Acoustic variability:
E.g. variations in duration or formant frequencies across instances (phones) of the same phoneme

- □ Inherent variability
- □ Effects of context -> "co-articulation"

P. 40, 8b

Coarticulation is the phenomenon of changes in articulation and acoustics of a phoneme due to its phonetic context.

I.e. gestures for successive phonemes overlap in time => V.T. shapes are affected.

Forward (anticipatory) articulation (R-L)

too

see

Carry-over (L-R) articulation: look-ahead planning

tell

bell

Prosody: "Suprasegmentals"

phone segments

- Supra-Segmental Aspect related to complete utterance

  Segment parameters varied – Pitch, Intensity, Duration

- **Linguistic Functions**
  - Phrasing: segmenting long utterances into clauses
  - Sentence Mode: declarative, question, exclamation
  - Prominent Words

The panda eats shoots and leaves.

Raj is going home

- **Paralinguistic Functions**
  - Emotions

Re-Duplication