Localising and Understanding the Neural Systems for Processing Spoken Words

Matt Davis
MRC Cognition & Brain Sciences Unit
Cambridge, UK
“Words are things, I’m convinced... Someday we’ll be able to measure the power of words. I think they are things. I think they get on the walls, they get in your wallpaper, they get in your rugs, in your upholstery, in your clothes. And, finally, into you.”
Processing Spoken Words

Localising vs explaining spoken word recognition

Bayesian inference in speech perception

Predictive computations for word recognition
Box & Arrow Models of Word Recognition

**Proposed Model**

- **Heard Words**
  - Acoustic
  - Superior temporal gyri
  - Posterior fusiform and lingual gyr
- **Written Words**
  - Visual
  - Posterior superior temporal sulcus
  - Posterior inferior temporal/Mid-fusiform
- **Semantics**
  - Non-semantic
  - Phonological retrieval
  - Extrasyllabian temporo-parietal
  - Semantic
  - Phonological retrieval
- **Articulatory Planning**
  - Frontal operculum
  - Anterior insula
  - Motor output & Hearing spoken response
  - Sensory-motor cortices & Superior temporal gyri

**Spoken Word**

- Acoustic analysis
- Auditory Input Lexicon
- Cognitive System
- retrieving the phonology (P) of seen words from orthography (O) can either occur via direct links (O-P) or indirectly via semantics (O-S, S-P).

**Written Word**

- Orthographic analysis
- Orthographic Input Lexicon
- Cognitive System
- Subword level Orthographic to Phonological conversion
- Phonological Output Lexicon
- Orthographic Output Lexicon
- Graphemic Output Buffer

**Connecting Arrow Models**

**O-P and P-O**

**P-S and S-P**

**O-S and S-O**

**Subword level Acoustic to Phonological conversion**

**Morton (1969) Logogen Model**

Patterson & Shewell (1987)

Price (2000, *J Anatomy*)
Computational accounts of Lexical Processing

TRACE model of speech perception (McClelland & Elman, 1986)

Interaction Activation Model of letter perception (McClelland & Rumelhart, 1981)
Mapping Computational Accounts onto the Brain

Interactive Activation Model of Letter Perception
(McClelland & Rumelhart, 1981)
Mapping Computational Accounts onto the Brain

Interactive Activation Model of Letter Perception (McClelland & Rumelhart, 1981)

Taylor, Rastle, & Davis (2013)
Psychological Bulletin
Linking Computational Accounts to the Brain

1. **Engagement:**
   Stimuli that are represented by a region lead to greater neural activity (e.g. words > pseudowords)

2. **Effort:**
   Stimuli that are a good fit to representations lead to less effort during neural processing (e.g. low > high frequency words)

Taylor, Rastle, & Davis (2013)
Psychological Bulletin
fMRI Meta-analysis: Written Words vs Pseudowords

Taylor, Rastle, & Davis (2013)
Psychological Bulletin
fMRI Meta-analysis: Spoken Words vs Pseudowords

Davis & Gaskell (2013) Phil Trans Roy Soc B

Processing Spoken Words

Localising vs explaining spoken word recognition
Processing Spoken Words

Localising vs explaining spoken word recognition

Bayesian inference for spoken words

Predictive computations for word recognition and learning
Bayesian Inference in Speech Perception

- **Posterior**: How probable is each word given the sound heard.
- **Likelihood**: How probable is hearing that sound when that word is said?
- **Prior**: How probable was each word before hearing any sound?

\[
P(\text{Word} \mid \text{Sound}) = \frac{P(\text{Sound} \mid \text{Word}) \times P(\text{Word})}{P(\text{Sound})}
\]

**Marginal**: How probable is hearing that sound.

Shortlist B: Norris & McQueen (2008, *Psychological Review*)
Davis & Scharenborg (2016, “Speech perception by humans & machines”)
Speech perception by machines

Traditional ASR System
from: Davis & Scharenborg (2016, in Gaskell & Mirkovic: Speech Perception & Spoken Word Recognition)
Neural Implementations of Bayesian Inference

(a) Predictive coding
   - prediction
     - prediction error
       - input

(b) Probability coding
   - prediction (prior prob.)
     - posterior prob.
       - input (likelihood)

(c) Log probability coding
   - prediction (log-prior prob.)
     - log posterior prob.
       - input (log-likelihood)

(d) Direct variable coding
   - prediction
     - posterior prob.
       - log posterior prob.
         - input

Current Opinion in Neurobiology

Aitchison & Lengyel (2017, Current Opinion in Neurobiology)
Vocoded Speech
(Simulation of a Cochlear Implant)

The man read the newspaper at lunchtime

Figure from: Davis et al (2005, Journal of Experimental Psychology: General)
Prior knowledge enhances speech clarity

Rate clarity of 1/2/4/8/16-channel vocoded words
Paired with matching/neutral/mismatching text

Sohoglu, Peelle, Carlyon & Davis (2014, JEP:HPP)
Prior knowledge enhances speech clarity

Sohoglu, Peelle, Carlyon & Davis (2014, *JEP:HPP*)
Prior knowledge and perception of speech

Prior knowledge and perception of speech

- Sohoglu, Peelle, Carlyon & Davis (2012, *J. Neuroscience*)
Prior knowledge and perception of speech

- Sohoglu, Peelle, Carlyon & Davis (2012, *J. Neuroscience*)
Prior knowledge and perception of speech

Superior temporal gyrus (STG)

- Sohoglu, Peelle, Carlyon & Davis (2012, J. Neuroscience)
Predictive coding model of speech perception

![Diagram showing the relationship between heard speech, predicted speech, and prediction error. The diagram includes a graph showing the change in evoked response over time with arrows representing the contribution of sensory detail and prior knowledge to the prediction of speech.](image)
Processing Spoken Words

Localising vs explaining spoken word recognition

Bayesian inference in speech perception

Predictive computations for word recognition
Recognising spoken words
(Cohort & TRACE Models)

“………………”

Blog
Cathedral
Cleave
Cathartic
Coat
Please

activation

Marslen-Wilson & Tyler (1980, Phil Trans B)

TRACE:
McClelland & Elman (1986, Cog Psych)
Recognising spoken words
(Cohort & TRACE Models)

“c

Blog
Cathedral
Cathartic
Cleave
Fish
Coat
Please

activation

TRACE:
McClelland & Elman (1986, Cog Psych)
Recognising spoken words
(Cohort & TRACE Models)

“ca

Blog

Cathedral

Can

Cleave

Fish

Cathartic

Coat

Please

activation

TRACE:
McClelland & Elman (1986, Cog Psych)
Recognising spoken words
(Cohort & TRACE Models)

“cath

Marslen-Wilson & Tyler (1980, *Phil Trans B*)

TRACE:
McClelland & Elman (1986, *Cog Psych*)
Recognising spoken words
(Cohort & TRACE Models)

“cathe

Blog

Cathedral

Cathartic

Cleave

Can

Fish

Coat

Please

activation

TRACE:
McClelland & Elman (1986, Cog Psych)
Recognising spoken words
(Cohort & TRACE Models)

“cathedr

Marslen-Wilson & Tyler (1980, Phil Trans B)

TRACE:
McClelland & Elman (1986, Cog Psych)
Recognising spoken words
(Cohort & TRACE Models)

“cathedral”

Marslen-Wilson & Tyler (1980, *Phil Trans B*)

TRACE:
McClelland & Elman (1986, *Cog Psych*)
Recognising spoken words

![Graph showing fixation probabilities over time for different types of words.]

**TABLE 2**

<table>
<thead>
<tr>
<th>Competitor set</th>
<th>Condition</th>
<th>Trials</th>
<th>Target</th>
<th>Distractors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noncompetitor 1</td>
<td>unrelated</td>
<td>3</td>
<td>referent</td>
<td>unrelated</td>
</tr>
<tr>
<td>Noncompetitor 2</td>
<td>referent</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme Only 1</td>
<td>unrelated</td>
<td>3</td>
<td>referent</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme Only 2</td>
<td>referent</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Full Competitor</td>
<td>unrelated</td>
<td>12</td>
<td>referent</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort Only</td>
<td>referent</td>
<td>6</td>
<td>cohort</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort Only</td>
<td>cohort</td>
<td>6</td>
<td>referent</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme</td>
<td>unrelated</td>
<td>6</td>
<td>rhyme</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme</td>
<td>rhyme</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort</td>
<td>referent</td>
<td>6</td>
<td>cohort</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort</td>
<td>cohort</td>
<td>6</td>
<td>referent</td>
<td>unrelated</td>
</tr>
<tr>
<td>Referent</td>
<td>cohort</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Referent</td>
<td>referent</td>
<td>6</td>
<td>cohort</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme</td>
<td>referent</td>
<td>6</td>
<td>rhyme</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme</td>
<td>rhyme</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort</td>
<td>referent</td>
<td>6</td>
<td>cohort</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort</td>
<td>cohort</td>
<td>6</td>
<td>referent</td>
<td>unrelated</td>
</tr>
<tr>
<td>Referent</td>
<td>cohort</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Referent</td>
<td>referent</td>
<td>6</td>
<td>cohort</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme</td>
<td>referent</td>
<td>6</td>
<td>rhyme</td>
<td>unrelated</td>
</tr>
<tr>
<td>Rhyme</td>
<td>rhyme</td>
<td>6</td>
<td>unrelated</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort</td>
<td>referent</td>
<td>6</td>
<td>cohort</td>
<td>unrelated</td>
</tr>
<tr>
<td>Cohort</td>
<td>cohort</td>
<td>6</td>
<td>referent</td>
<td>unrelated</td>
</tr>
</tbody>
</table>

![Schematic of the grid with pictures.](image)

Alloppenna, Magnuson & Tanenhaus (1998, JML)
Recognising spoken words

![Graph showing fixation probability over time](image)

![Graph showing activation in TRACE](image)

Alloppenna Magnuson & Tanenhaus (1998, JML)
Competitive vs Predictive Selection

TRACE
(McClelland & Elman, 1986)

Words

Phonemes

Acoustic Features

Competitive Selection

Interactive Activation

Marslen-Wilson (1984, Attention & Performance X)
Marslen-Wilson (1987, Cognition)
Sharpening vs Predictive Coding

TRACE (McClelland & Elman, 1986)

Competitive Selection

Interactive Activation

Words

Phonemes

Acoustic Features

Predictive Coding

Prediction Error

Top Down Predictions

Words

Phonemes

Acoustic Features

Davis & Sohoglu (2020 Cog Neurosci 6, MIT Press)
https://psyarxiv.com/qc4u6/
Predictive coding and word recognition

After hearing: /kæptɪn/

Probability(“Captain”) = 0.9
Probability(“Captive”) = 0.1

P(/n/) ~ 1
P(/v/) ~ 0

P(/n/) = 0.9
P(/v/) = 0.1

Δ “Captain” = +0.1
Δ “Captive” = -0.1

Gagnepain, Henson & Davis (2012, Current Biology)
Predictive coding and word recognition

After hearing: /kæptɪv/

Probability(“Captain”) = 0.9
Probability(“Captive”) = 0.1

P(/n/) ~ 0
P(/v/) ~ 1

Sounds

Heard Speech

- Predicted Speech

Prediction Error

Δ “Captain” = -0.9
Δ “Captive” = +0.9

Signals lexical match and mismatch

Gagnepain, Henson & Davis (2012, Current Biology)
Speech predictions change with learning

Hygiene

Hygiene

Hygiene \rightarrow \text{Expected Sounds}

Prediction Error

Delays word recognition (Monsell & Hirsh, 1998)

Predictive Selection (cf. Davis & Sohoglu, 2020)

Heard Sounds

Localise lexical prediction error to STG regions

Hijack

Hijous

Speech predictions change with learning.
Competitive Selection vs Predictive Selection

**Input:**
- /h/
- /ai/
- /dʒ/

**DP:**
- /ə/
- /s/

**Lexical Uncertainty (Entropy):**
- Habit
- Hack
- ... health
- Help
- ... hijack
- ... hobby
- ... hygiene

Wang, Sohoglu, Gilbert, Henson & Davis (2021, J Neurosci)
Competitive Selection vs Predictive Selection

Input: /h/ /ai/ /dʒ/ /n/

Prediction:

Segment Surprisal (Prediction Error):

Wang, Sohoglu, Gilbert, Henson & Davis (2021, J Neurosci)
Competitive Selection vs Predictive Selection

Input: /h/ /ai/ /dʒ/ /k/
Prediction:
/æ/ /e/ /i:/: /k/ /n/

Segment Surprisal (Prediction Error):

Wang, Sohoglu, Gilbert, Henson & Davis (2021, J Neurosci)
Competitive Selection vs Predictive Selection

Input: /h/ /ai/ /dʒ/ /ə/ /s/

Prediction:

Segment Surprisal (Prediction Error):

Wang, Sohoglu, Gilbert, Henson & Davis (2021, *J Neurosci*)
Competitor Priming for Word Pairs (Behaviour)

Wang, Sohoglu, Gilbert, Henson & Davis (2021, *J Neurosci*)
Competitor Priming for Word Pairs overlaps with Pseudo > Word (MEG)

STG Source of Lexicality Effect

Wang, Sohoglu, Gilbert, Henson & Davis (2021, J Neurosci)
Competitor Priming for Word Pairs overlaps with Pseudo > Word (MEG)

STG Source of Lexicality Effect

Trial-wise correlation of neural and behavioural priming

Wang, Sohoglu, Gilbert, Henson & Davis (2021, *J Neurosci*)
Localising vs explaining spoken word recognition

Bayesian inference in speech perception

Predictive computations for word recognition

\[
P(\text{Word | Sound}) = \frac{P(\text{Sound | Word}) \times P(\text{Word})}{P(\text{Sound})}
\]
References from question period:

- **Word learning and overnight consolidation:**
  - **Behaviour:** Dumay & Gaskell (2007, Psych Science) [https://doi.org/10.1111/j.1467-9280.2007.01845.x](https://doi.org/10.1111/j.1467-9280.2007.01845.x)
  - **fMRI:** Davis et al (2009, JoCN) [https://doi.org/10.1162/jocn.2009.21059](https://doi.org/10.1162/jocn.2009.21059)
  - **MEG:** Gagnepain et al (2012, Current Biology) [https://doi.org/10.1016/j.cub.2012.02.015](https://doi.org/10.1016/j.cub.2012.02.015)
  - **Theory / Review:** Davis & Gaskell (2009, Phil Trans Roy Soc B) [https://doi.org/10.1098/rstb.2009.0111](https://doi.org/10.1098/rstb.2009.0111)

- **Responses to pseudowords of different lengths:**
  - **EEG:** O’Rourke & Holcomb (2002, Biological Psychology) [https://doi.org/10.1016/s0301-0511(02)00045-5](https://doi.org/10.1016/s0301-0511(02)00045-5)
  - **fMRI:** Zhuang et al (2014, Cerebral Cortex) [https://doi.org/10.1093/cercor/bhs366](https://doi.org/10.1093/cercor/bhs366)

- **RSA & Decoding tests TRACE/Sharpening vs Prediction error**
  - **fMRI:** Blank & Davis (2016, PLoS Biology) [https://doi.org/10.1371/journal.pbio.1002577](https://doi.org/10.1371/journal.pbio.1002577)
  - **MEG:** Sohoglu & Davis (2020, eLife) [https://doi.org/10.7554/eLife.58077](https://doi.org/10.7554/eLife.58077)