Speech perception: As abstract as it needs to be

James M. McQueen

Radboud University, Nijmegen

Donders Institute for Brain, Cognition & Behaviour, Centre for Cognition

& Max Planck Institute for Psycholinguistics

j.mcqueen@donders.ru.nl

Phonetic diversity

Each utterance we hear is unique

- Different words, talkers, contexts

But listeners can cope

- Immediate recognition of novel utterances

"Kartoffelpüree mit Bratwurst und Sauerkraut"

Coping with variability

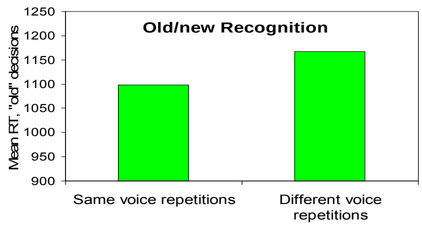
Hybrid model:

1. Storage of abstract representations of:
   - Segments
   - Suprasegmental structures
   - Words
2. Storage of episodic memories
   - Talker-specific, indexical and situational details
   - So (groups of) talkers can be understood better in the future

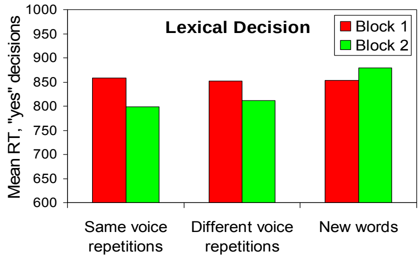
Perception is as abstract as it needs to be

Lexical memory vs. episodic memory

Voice-specificity effects (repetition priming) in recognition memory task



But no voice effects in lexical decision



Voice-specific detail is stored in long-term episodic memory, but not in the mental lexicon

Luce & Lyons (1998)

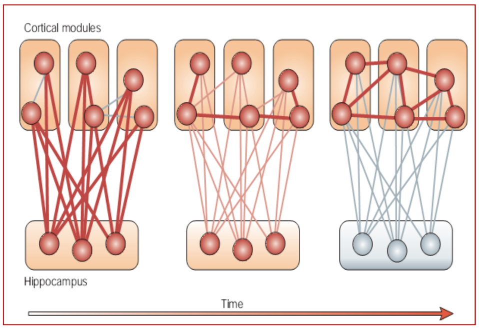
Complementary Learning Systems model

Episodic memory

– Fast, initial hippocampal (and medial-temporal) learning

Semantic memory

– Slower, later neocortical learning through consolidation

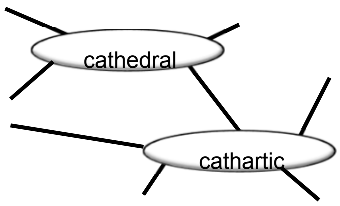


McClelland et al. (1995)

Sleep and memory consolidation in word learning

hippocampus
episodic, temporary

cortex
integrated, stable



A word has become lexicalized when it starts to compete with other words (Gaskell & Dumay, 2003)

Competition with new words emerges after (sleep-enhanced) consolidation (Dumay & Gaskell, 2007)

Testing consolidation in word learning

Training: Phoneme monitoring

x 36 blocks

n

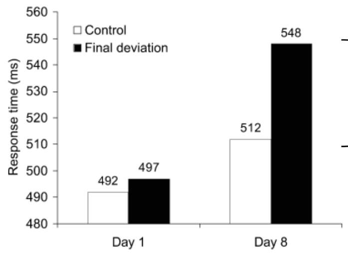
cathedruke

alcohin

Test: Pause detection

cathedral

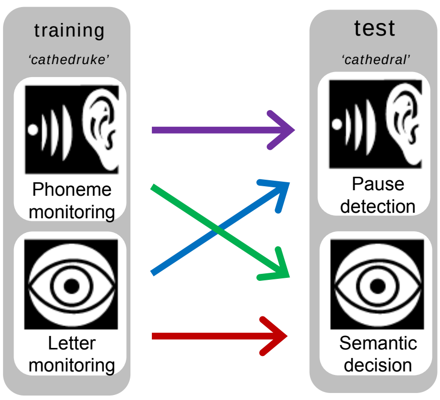
parach...ute



Competition effect

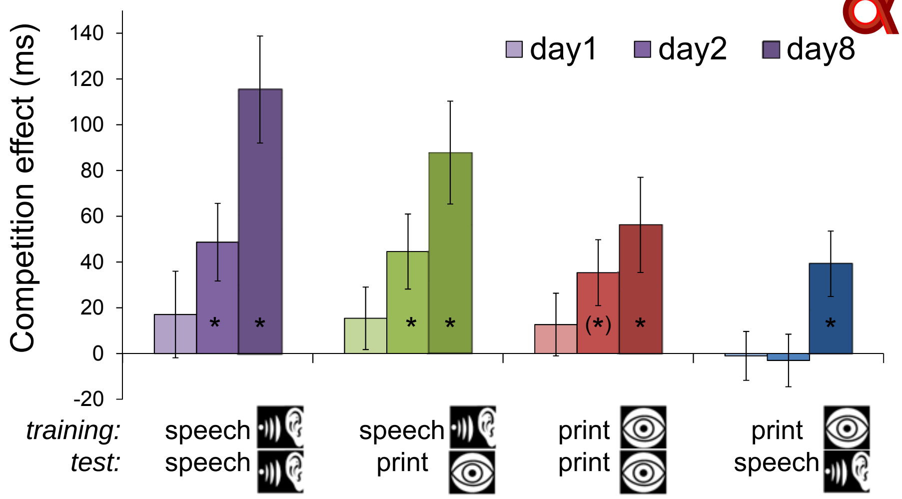
Gaskell & Dumay (2003)

Cross-modal generalization



day 1	day 2	day 8
training		
competition task	competition task	competition task
free recall	free recall	free recall
recognition	recognition	recognition
	training	

Bakker, Takashima, van Hell, Janzen & McQueen (2014)



Abstraction beyond episodic experience:
Competition in print from words never seen before, and in speech from words never heard before

Bakker, Takashima, van Hell, Janzen & McQueen (2014)

Coping with variability

Hybrid storage

1. Abstract linguistic representations
2. Episodic memories

Abstraction

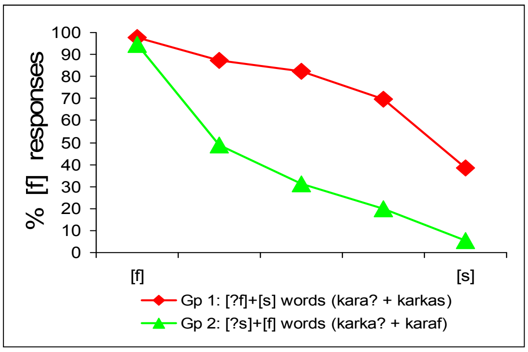
- Linking variable forms to meanings

Adaptation

- Tuning in to variability

Lexical retuning of phonetic categories

- Part 1: Lexical decision
  - Gp 1: 20 ambiguous [f]-final & 20 natural [s]-final words (e.g. *kara?* & *karkas*)
  - Gp 2: 20 ambiguous [s]-final & 20 natural [f]-final words (e.g. *karka?* & *karaf*)
- Part 2: Phonetic categorisation
  - Identify sounds on [ɛf] -- [ɛ?] -- [ɛs] continuum



Norris, McQueen & Cutler (2003)

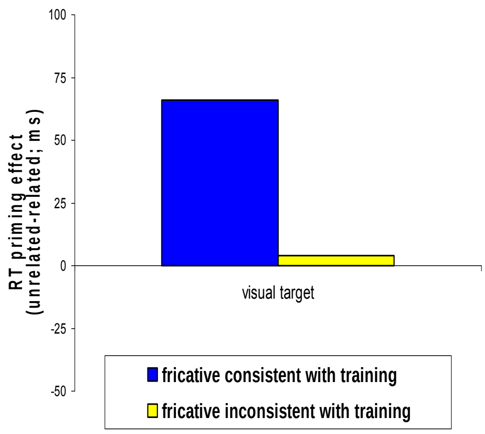
Generalization to new words

Cross-modal identity priming with minimal pairs such as doof/doos ("deaf"/"box")

Responses were faster after related than after unrelated primes, but only when the target's final sound was consistent with the lexically-biased training

Gp1: [do:?]-doof << [krop]-doof
Gp2: [do:?]-doos << [krop]-doos

Gp1 hear [do:?] as *doof*,
Gp2 hear [do:?] as *doos*



McQueen, Cutler & Norris (2006)
Sjerps & McQueen (2010)

Lexically-guided retuning of segment perception

Retuning helps listeners cope with speech variability

- It generalizes to other words, and:
  - Can be talker specific (Eisner & McQueen, 2005)
  - Is stable over time (Eisner & McQueen, 2006)
  - Is possible in a second language (Mitterer & McQueen, 2009)
  - Is transferable across positions (Jesse & McQueen, 2011)

Generalization of learning across the vocabulary depends on prelexical abstraction about segments

These abstractions play a **functional role**

- So learning paradigms can reveal the units of perception

What are the prelexical units of perception?

Position-invariant phonemes or position-specific allophones?

Test with allophonically variable Dutch liquids:
/r/ is approximant or trill; /l/ is light or dark

Part 1: auditory lexical decision:

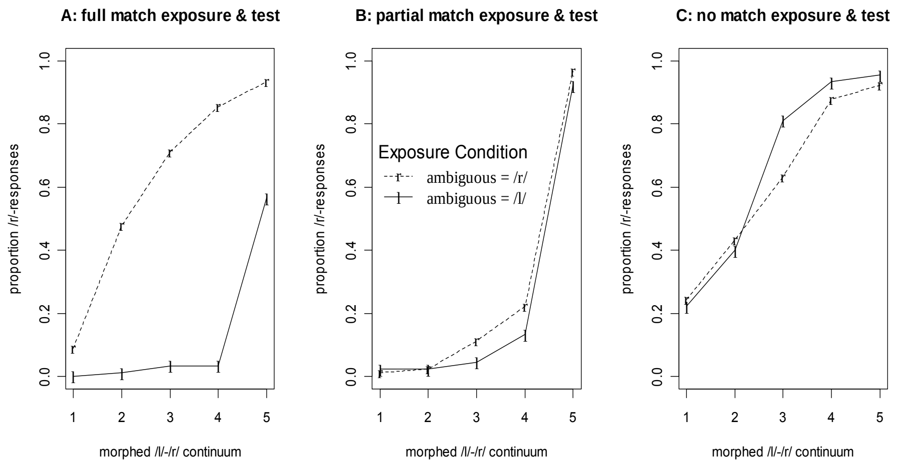
- [?] midway between approximant [ʴ] and dark [ɫ]
- Gp1: learning [?] is /r/ or approximant [ʴ]? (*bakke?* + *appel*)
- Gp2: learning [?] is /l/ or dark [ɫ]? (*bakker* + *appe?*)

Part 2: categorization of nonword-nonword continuum:

A. *kwipter-kwiptel*: in coda, approximant [ɹ] to dark [ɫ]
B. *kwipter-kwiptel*: in coda, trill [r] to dark [ɫ]
C. *repaas-lepaas*: in onset, trill [r] to light [l]

- If retuning is phonemic, effect should be seen on all 3 continua
- If retuning is allophonic, effect only when there is full match between exposure and test sounds

Retuning is about allophones, not phonemes



Retuning in the exposure phase on approximant [ɹ] and dark [ɫ] applied at test only to these allophones, not to all /r/ and all /l/

Mitterer, Scharenborg & McQueen (2013)

Or phonemes after all?

Selective adaptation

Allophonically variable English stops

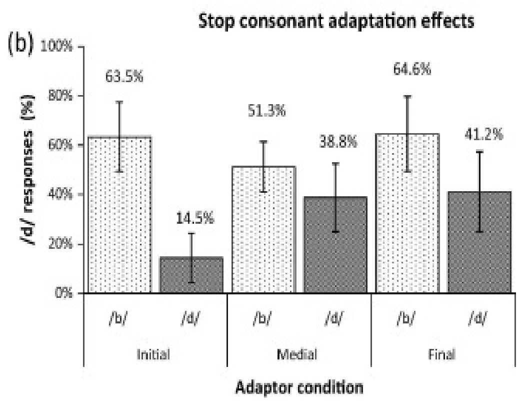
Adaptation: 25 words with unambiguous /b/ (e.g. "*bail*") or /d/ (e.g. "*desk*")

Test: categorise "*?ump*" (*bump* or *dump*)

Adaptors in initial, medial or final position

Test stimuli always in initial position

6 cycles of pairs of adaptation+test sessions



- Adaptation generalized across positions
- "Spoken word identification involves accessing position invariant phoneme representations"

Bowers, Kazanina & Andermane (2016)

Selective adaptation

Some problems with Bowers et al. (2016)

- If units are phonemes, why the interaction with position?
- Were the stops acoustically very different across position?
  - 11/25 final /b/'s and 25/25 final /d/'s had release bursts

Stronger tests:

- Dutch liquids
- German fricatives

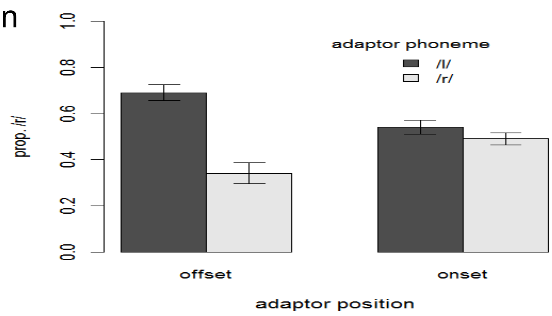
Dutch liquids

Adaptation:

Offset [ɫ]: e.g. appel Offset [ɹ]: e.g. bakker
Onset [l]: e.g. leiding Onset trill [r]: e.g. rente

Test: Offset [wɪmpəɫ] – [wɪmpəɹ] (wimpel – wimper) continuum

Evidence of adaptation only if adaptors and test stimuli shared allophones



Mitterer, Reinisch & McQueen (subm.)

German fricatives

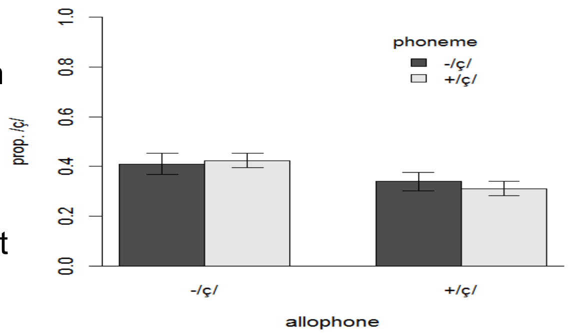
Adaptation:

Adaptor overlap	Orthography	Underlying	Surface
+phonemic, +allophonic	friedlich	/fridlɪç/	[fridlɪç]
+phonemic, -allophonic	flach	/flaç/	[flax]
-phonemic, +allophonic	König	/kø:nɪg/	[kø:nɪç]
-phonemic, -allophonic	Auge	/aʊgə/	[aʊgə]

Test: [kɪɐçə] – [kɪɐʃə] (Kirche – Kirsche) continuum

Evidence of adaptation only with allophonic overlap

Even when the phonemes are different



Mitterer, Reinisch & McQueen (subm.)

What are the prelexical units of perception?

Lexically-guided retuning

- Position-specific allophones

Selective adaptation

- Position-specific allophones
- **Bowers et al.**: Apparent phonemic effects due to acoustic overlap
- **Mitterer et al.**: Evidence for the null hypothesis for phonemes:
  - Dutch liquids: Bayes Factor = 0.21
  - German fricatives: Bayes Factor = 0.13

Why allophones?

- Tuning in to speech allophonically helps the listener; tuning in phonemically does not

With thanks to:

Iske Bakker, Anne Cutler, Frank Eisner,
Janet van Hell, Gabriele Janzen,
Alexandra Jesse, **Holger Mitterer**,
Dennis Norris, **Eva Reinisch**,
Odette Scharenborg, Matthias Sjerps
& Atsuko Takashima

NWO

DONDERS INSTITUTE

Conclusions

Hybrid storage
- Abstract linguistic representations
- Episodic memories

Abstraction
- Prelexical and lexical
- Segmental (and suprasegmental)

Adaptation
- Tuning in to variability about allophones helps listener cope with phonetic diversity
- Perception is as abstract as it needs to be

Radboud University Radboudumc