Speech perception and listener adjustments

Frank Eisner

Donders Centre for Cognition
Radboud University Nijmegen

2/10/2013

Does information about the speaker help listeners extract meaning?

Do individual differences matter?

This talk

I. Models of speech perception

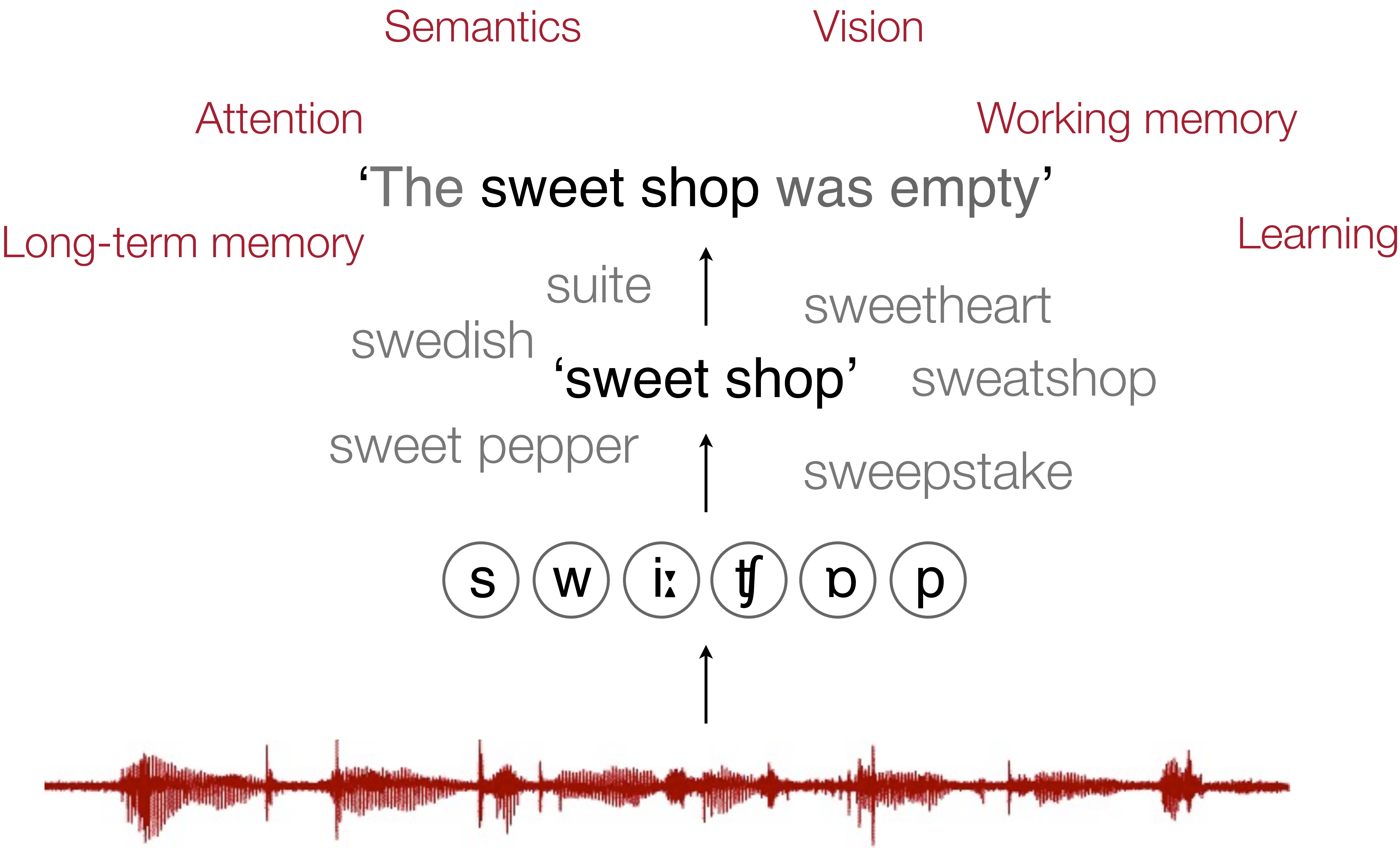
II. Perceptual learning

Adjusting pre-lexical categories

Adjusting to global signal degradation

III. Implications for models and applications

Abstractionist view of speech perception



The TRACE model, 1986

COGNITIVE PSYCHOLOGY 18, 1–86 (1986)

The TRACE Model of Speech Perception

JAMES L. MCCLELLAND

Carnegie–Mellon University

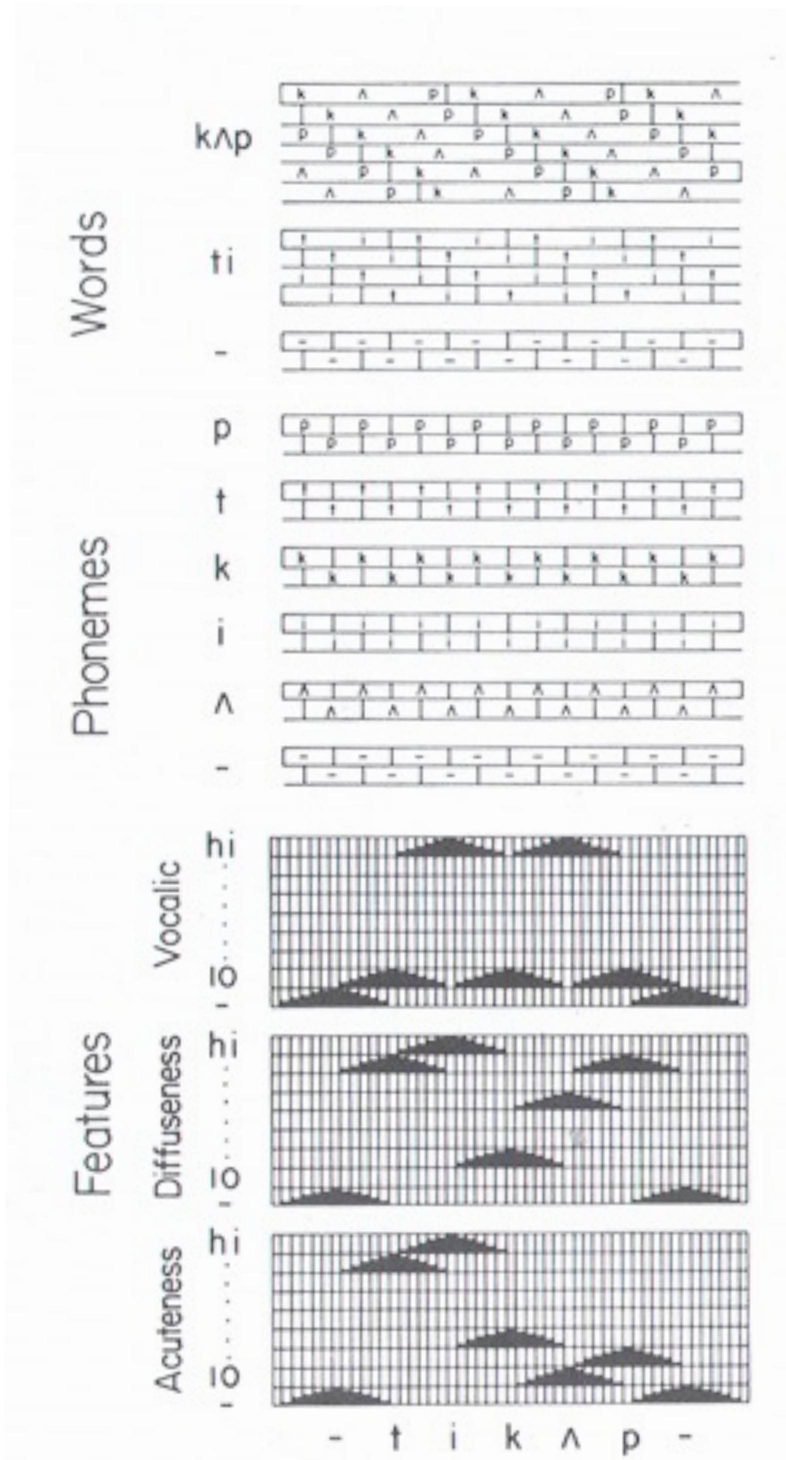
AND

JEFFREY L. ELMAN

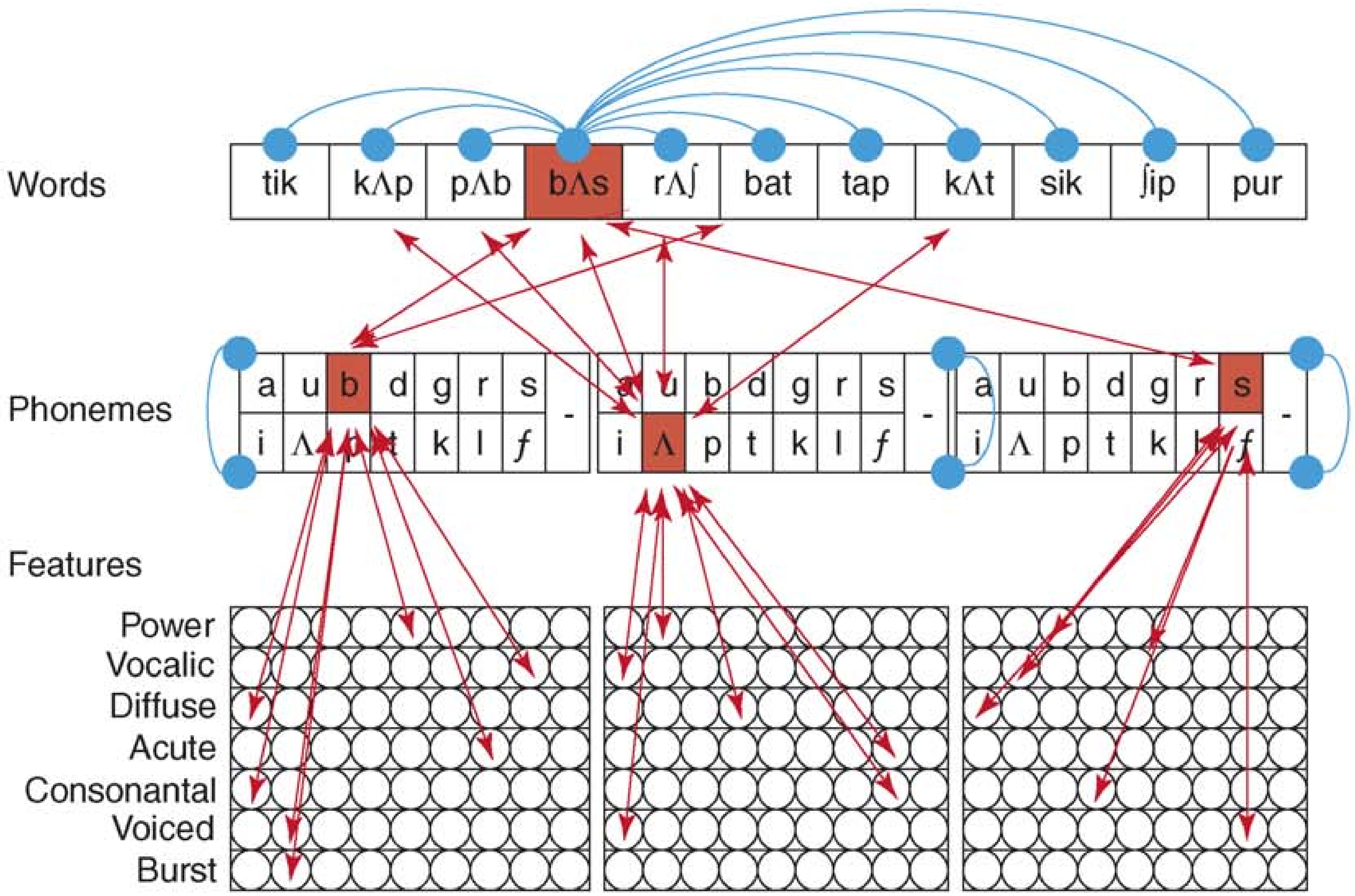
University of California, San Diego

We describe a model called the TRACE model of speech perception. The model is based on the principles of interactive activation. Information processing takes place through the excitatory and inhibitory interactions of a large number of simple processing units, each working continuously to update its own activation on the basis of the activations of other units to which it is connected. The model is called the TRACE model because the network of units forms a dynamic processing structure called "the Trace," which serves at once as the perceptual processing mechanism and as the system's working memory. The model is instantiated in two simulation programs. TRACE I, described in detail elsewhere, deals with short segments of real speech, and suggests a mechanism for coping with the fact that the cues to the identity of phonemes vary as a function of context. TRACE II, the focus of this article, simulates a large number of empirical findings on the perception of phonemes and words and on the interactions of phoneme and word perception. At the phoneme level, TRACE II simulates the influence of lexical information on the identification of phonemes and accounts for the fact that lexical effects are found under certain conditions but not others. The model also shows how knowledge of phonological constraints can be embodied in particular lexical items but can still be used to influence processing of novel, nonword utterances. The model also exhibits categorical perception and

The work reported here was supported in part by a contract from the Office of Naval Research (N-00014-82-C-0374), in part by a grant from the National Science Foundation (BNS-79-24062), and in part by a Research Scientists Career Development Award to the first author from the National Institute of Mental Health (5-K01-MH00385). We thank Dr. Joanne Miller for a very useful discussion which inspired us to write this article in its present form. David Pisoni was extremely helpful in making us deal more fully with several important issues, and in alerting us to a large number of useful papers in the literature. We also thank David Rumelhart for useful discussions during the development of the basic architecture of TRACE and Eileen Conway, Mark Johnson, Dave Pare, and Paul Smith for their assistance in programing and graphics. Send requests for reprints to James L. McClelland, Department of Psychology, Carnegie–Mellon University, Schenley Park, Pittsburgh, PA 15213.



The TRACE model, 2006



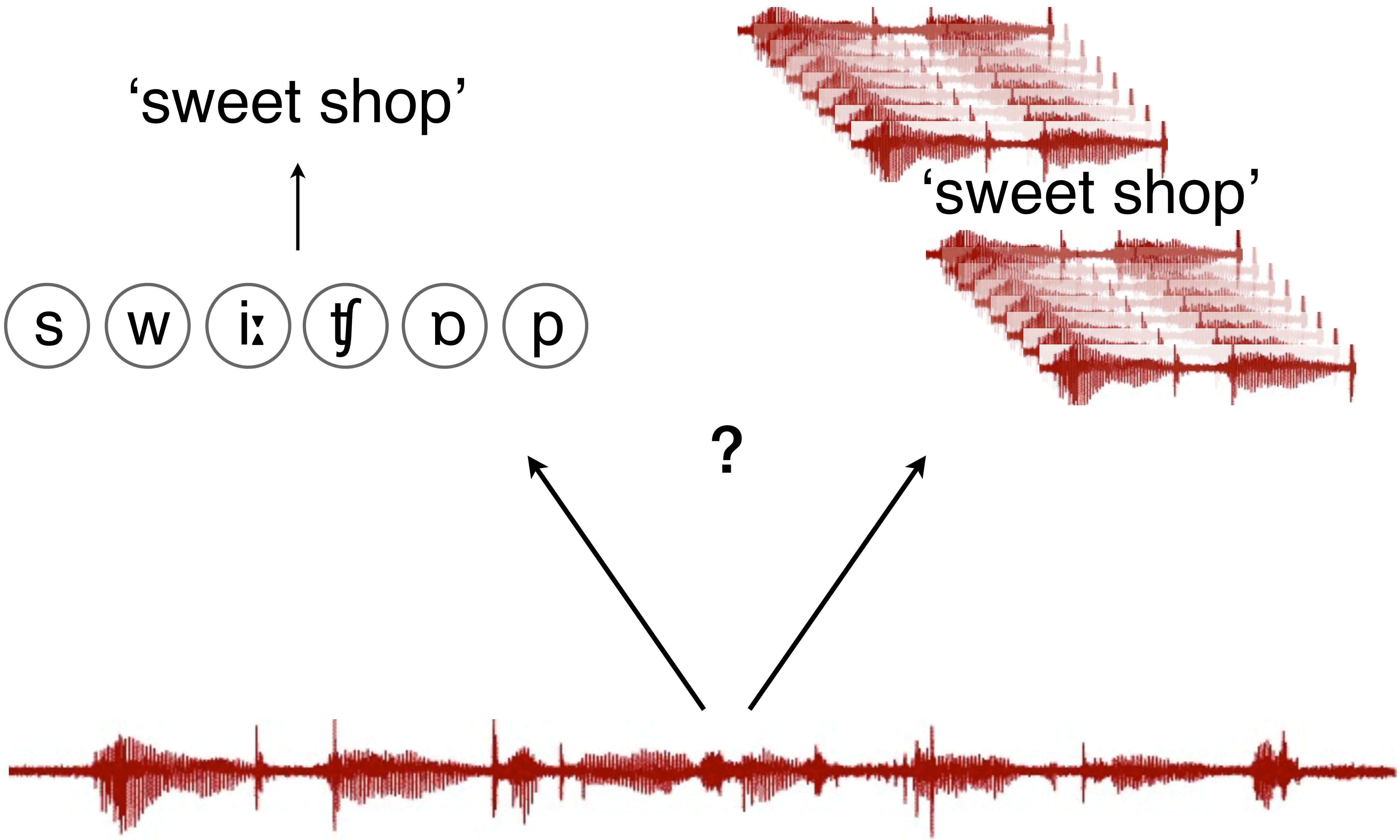
TRACE's architecture:

Red: Bi-directional excitatory connections.
Mutually consistent elements at adjacent levels support each other through excitation.

Blue: Units within a layer compete through inhibitory connections.

The episodic view

Which is right?



Perceptual learning in speech

"Perceptual learning [...] refers to an increase in the ability to extract information from the environment, as a result of experience and practice with stimulation coming from it."

(Gibson 1969)

This talk

I. Models of speech perception

II. Perceptual learning

Adjusting pre-lexical categories

Adjusting to global signal degradation

III. Implications for models and applications

Adjusting pre-lexical categories

Listeners can use lexical information to adjust their perception of an ambiguous sound:

An ambiguous fricative sound /sf?/ is perceived

as /s/ in a context like *albatro–*

as /f/ in a context like *paragra–*

After repeated exposure to the sound in such lexically-biased contexts, listeners adjust their category boundary between /s/ and /f/ in a way that is consistent with the lexical context.

Norris, McQueen & Cutler (2003) Cognitive Psychology

Adjusting pre-lexical categories

Pretest:
f/s categorization

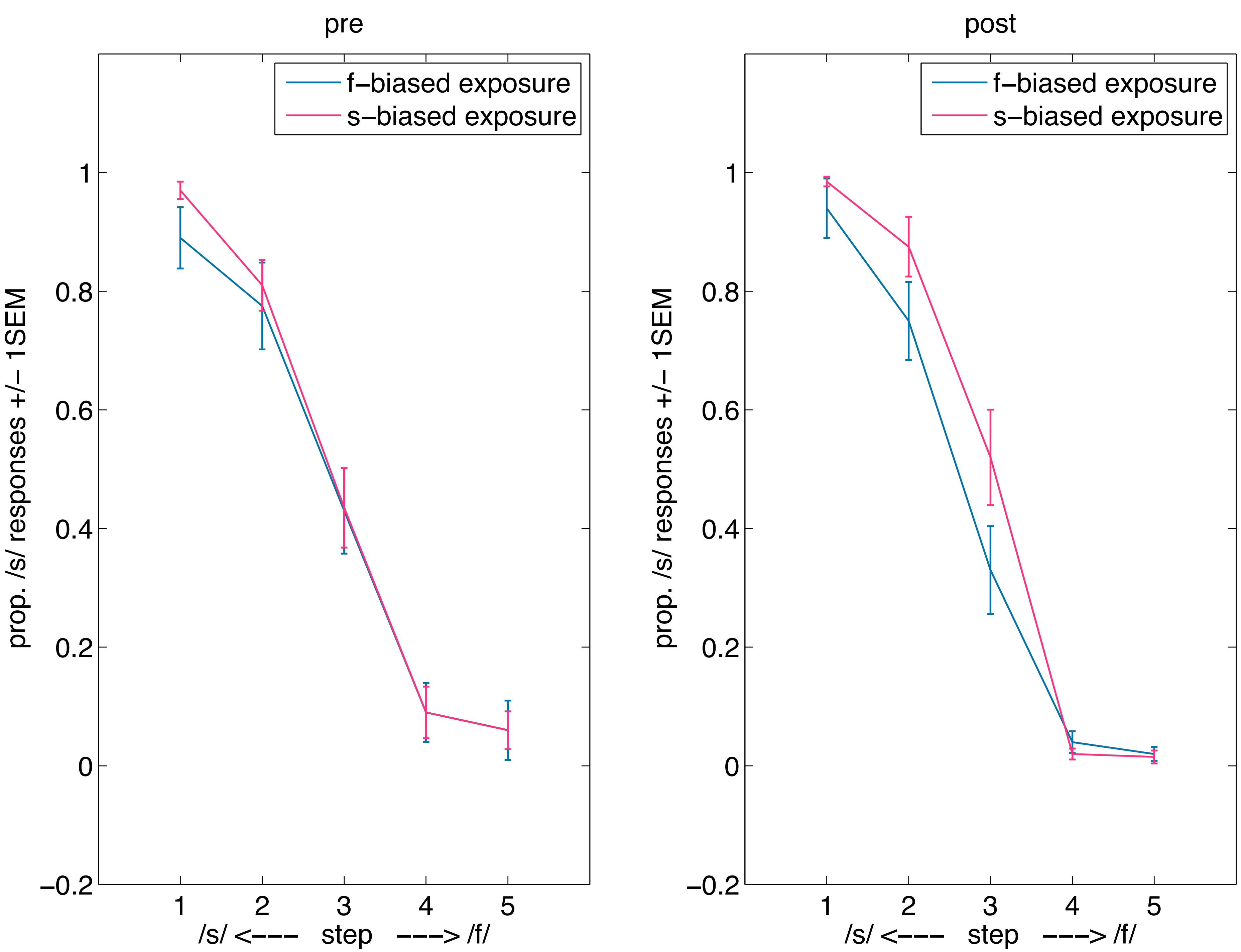
Exposure:
Ambiguous **/sf?/** sound embedded in continuous speech

s-biased
"De uitgang**s**positie van on**s** land i**s** en blijft relatief po**s**itief"

f-biased
"De uitgangspositie **v**an ons land is en blij**f**t relatie**f** positie**f**"

Posttest:
f/s categorization

Adjusting pre-lexical categories



Eisner, Iltzsche & Weber (in prep)

Effects of long-term familiarity with a speaker?

Perceptual learning occurred regardless of familiarity with the speaker.

Adjusting pre-lexical categories

Listeners use lexical information to adjust their perception of an ambiguous sound. — Norris et al. 2003

This type of learning

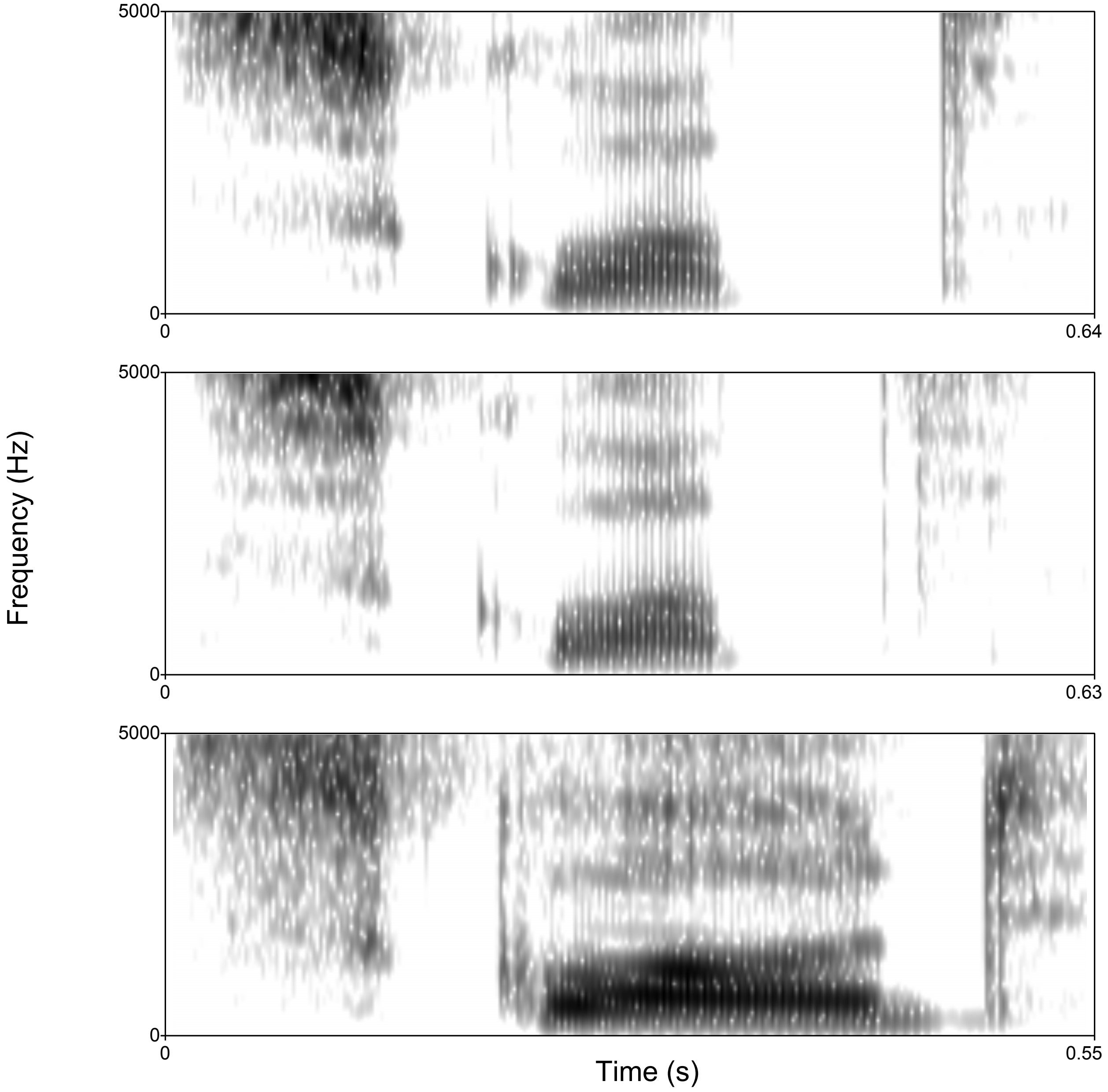
- does not require explicit attention — McQueen et al. 2006
- can be specific for a particular talker — Eisner & McQueen 2005

—or not, depending on the type of phoneme — Kraljic & Samuel 2006

- remains stable for at least 12 hours — Eisner & McQueen 2006
- can be modulated by context — Kraljic et al. 2008
- generalizes across the lexicon — McQueen et al. 2006
- also works with visual information from the face — Bertelson et al. 2003
- also works using phonotactic constraints — Cutler et al. 2008

Encoding of position?

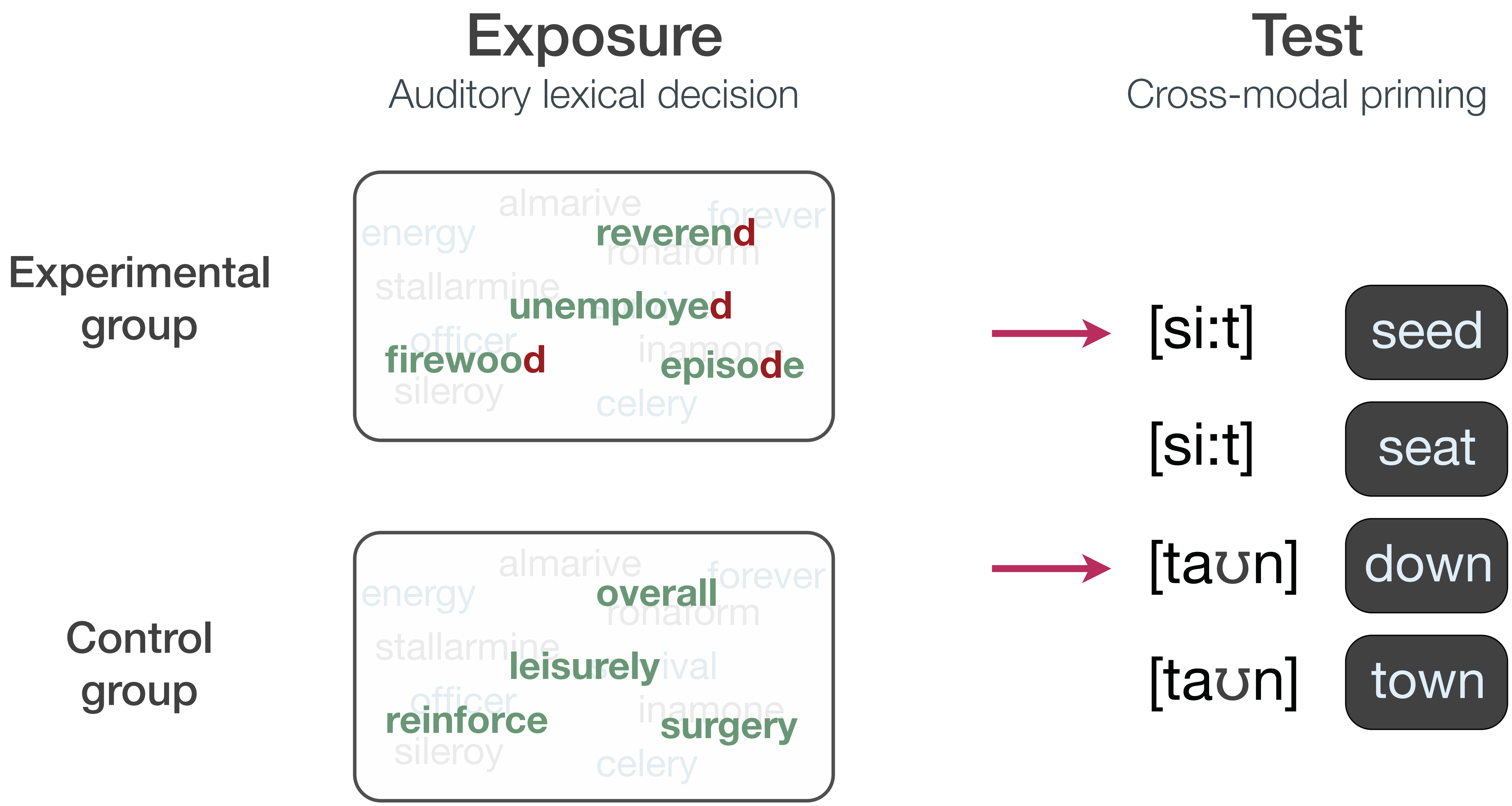
Word-final devoicing in Dutch-accented English



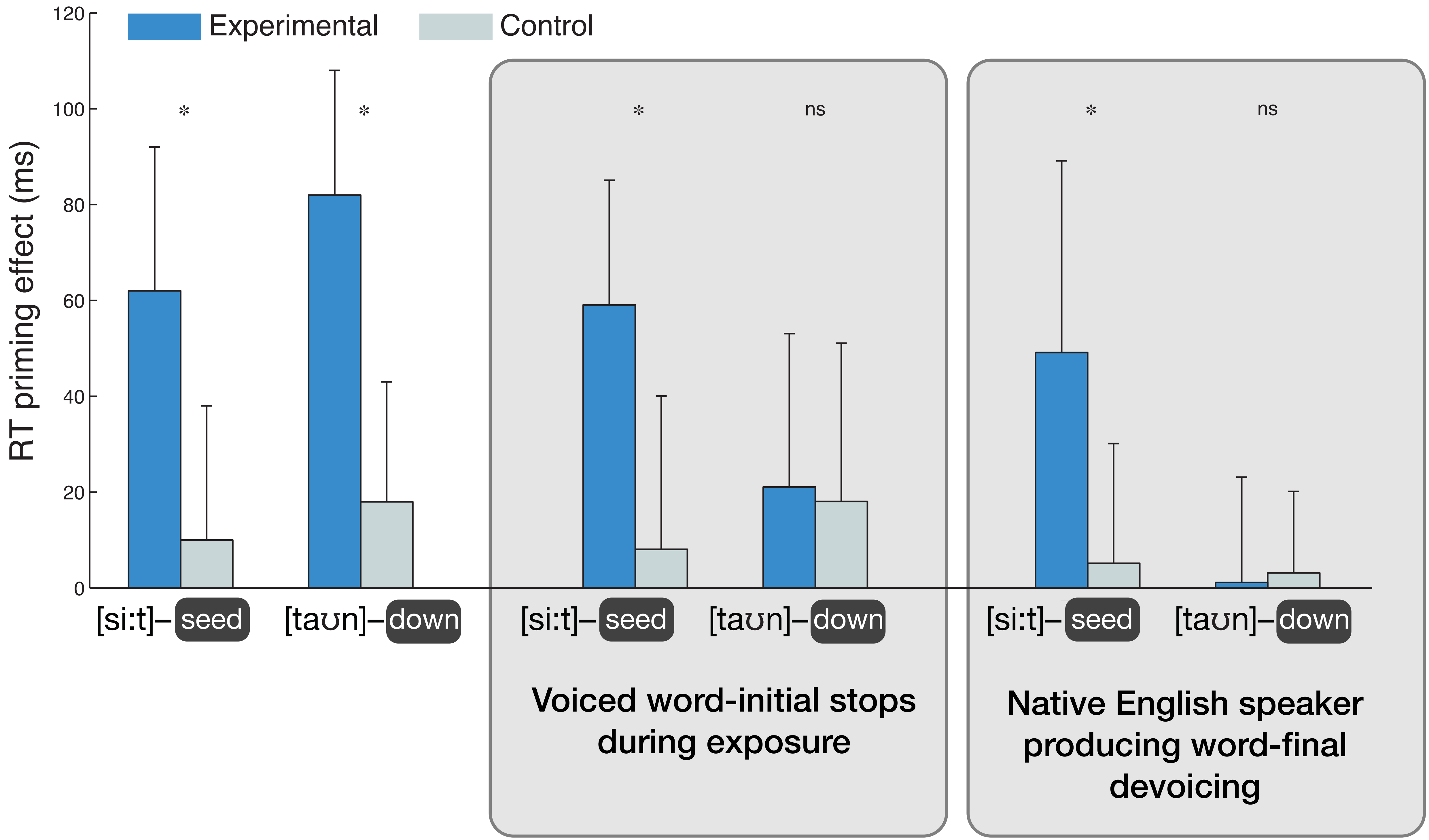
squad (NL)

squat (NL)

squad (E)



Exposure
Auditory lexical decision
Test
Cross-modal priming
Experimental group
almarive
energy
forever
reverend
ronaform
stallarmine
unemployed
officer
firewood
inamone
episode
sileroy
celery
Control group
almarive
energy
forever
overall
ronaform
stallarmine
leisurely
officer
reinforce
inamone
surgery
sileroy
celery
[si:t]
seed
[si:t]
seat
[taʊn]
down
[taʊn]
town



Eisner, Melinger & Weber (2013)

- Perceptual learning generalized to word-initial position when listeners had not heard normally-voiced initial stops.

- However, this generalization was constrained when listeners had exposure to normally-voiced initial stops, and when the speaker had no general foreign accent.

- This suggests that there can be sensitivity to the position in which the critical phoneme occurs.

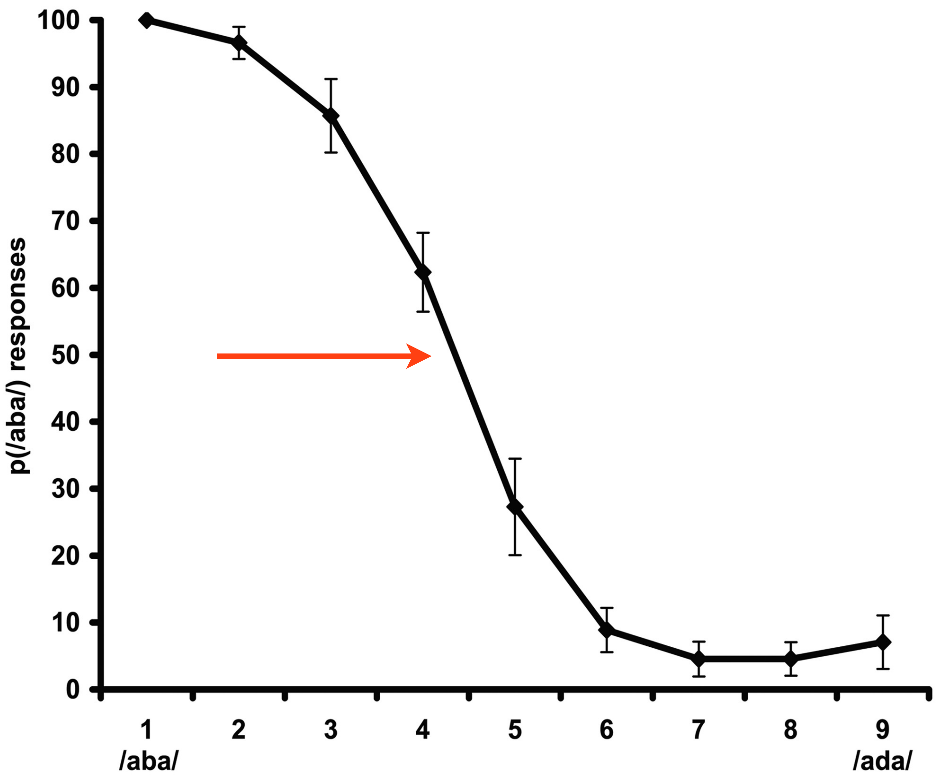
- Learning generalised from exposure words to a novel set of test words.

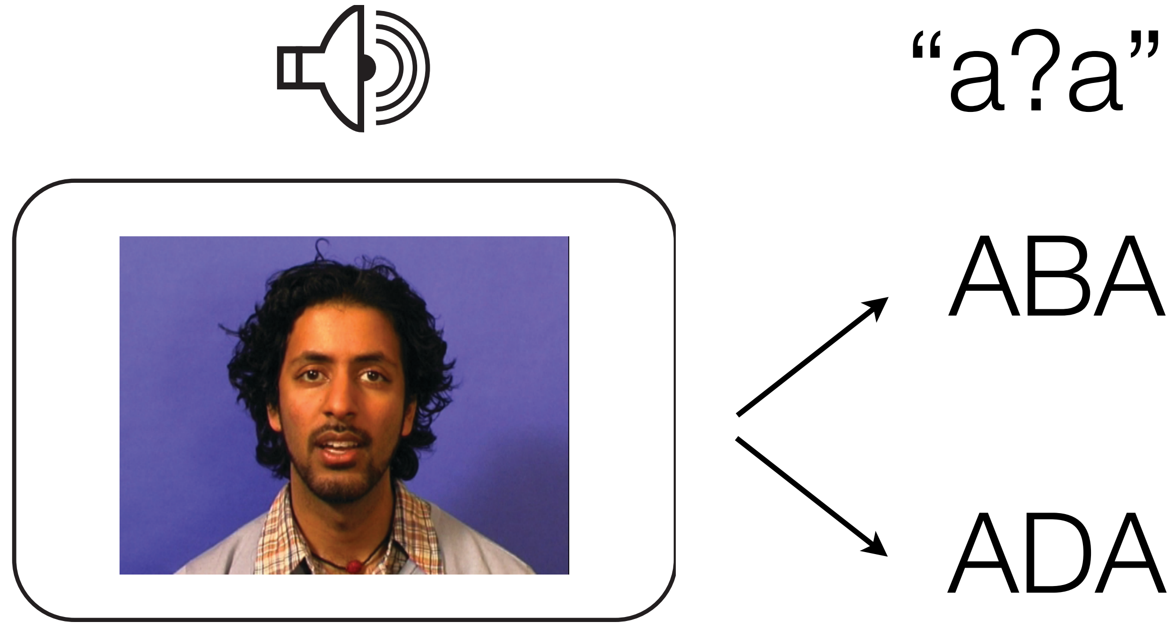
The Journal of Neuroscience, February 2, 2011 • 31(5):1715–1720 • 1715

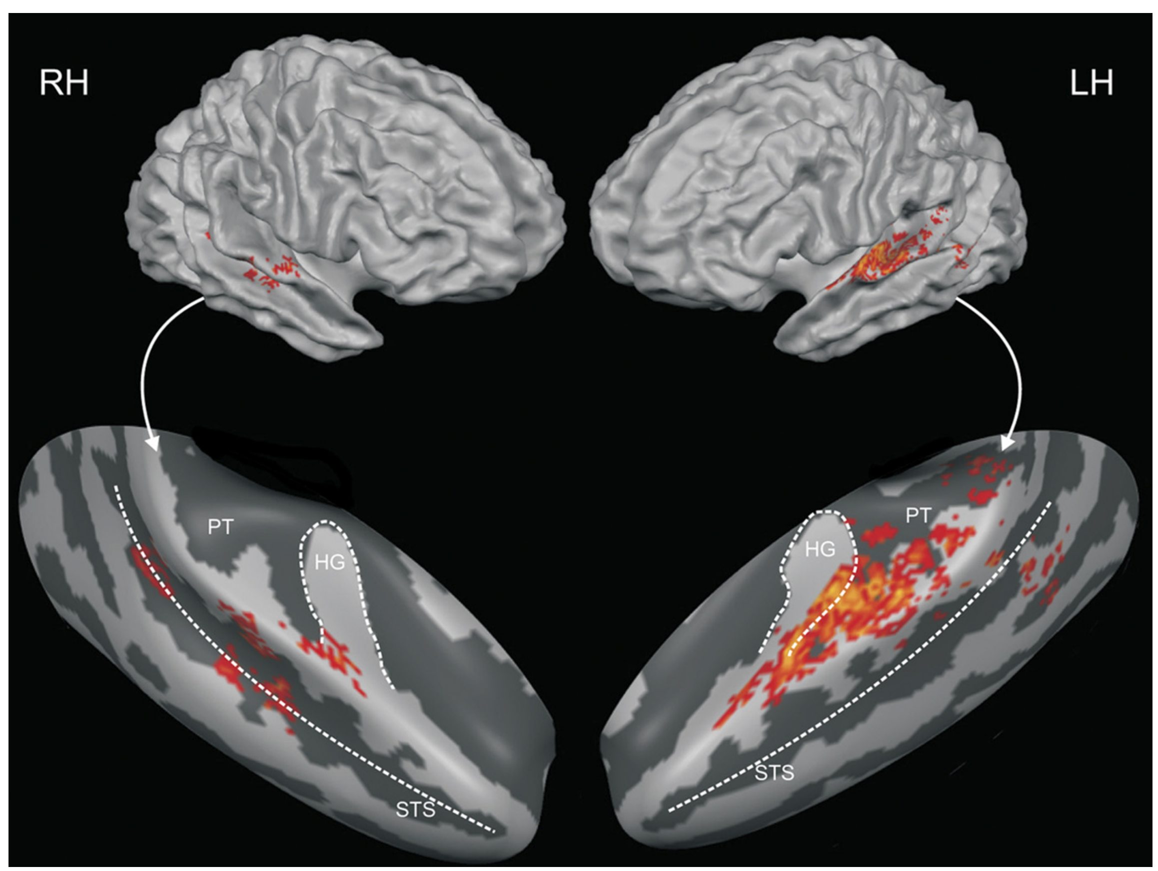
Behavioral/Systems/Cognitive

Auditory Cortex Encodes the Perceptual Interpretation of Ambiguous Sound

Niclas Kilian-Hütten,[1,2] Giancarlo Valente,[1,2] Jean Vroomen,[3] and Elia Formisano[1,2]
[1]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, and [2]Maastricht Brain Imaging Center, Maastricht University, 6200 MD Maastricht, The Netherlands, and [3]Department of Psychology, Tilburg University, 5000 LE Tilburg, The Netherlands







Kilian-Hütten, Valente, Vroomen, & Formisano (2011) Journal of Neuroscience.

“We propose that what is reflected in the reliance on early networks in this case is [...] a perceptual bias that is stored in these regions.

This bias is responsible for the behavioral (perceptual) effect and is installed by the cross-modal recalibration mechanism.

Its origin may lie within higher-order areas involved in the integration of audiovisual speech signals [...]. The information is then fed back from there to early auditory areas.

Here, [...] the perceptual bias is stored and sensory input is transformed into more abstract entities or auditory objects. These abstract entities may be considered as the building blocks of further linguistic and vocal processing.”

.Kilian-Hütten, Valente, Vroomen, & Formisano (2011) Journal of Neuroscience.

This talk

I. Models of speech perception

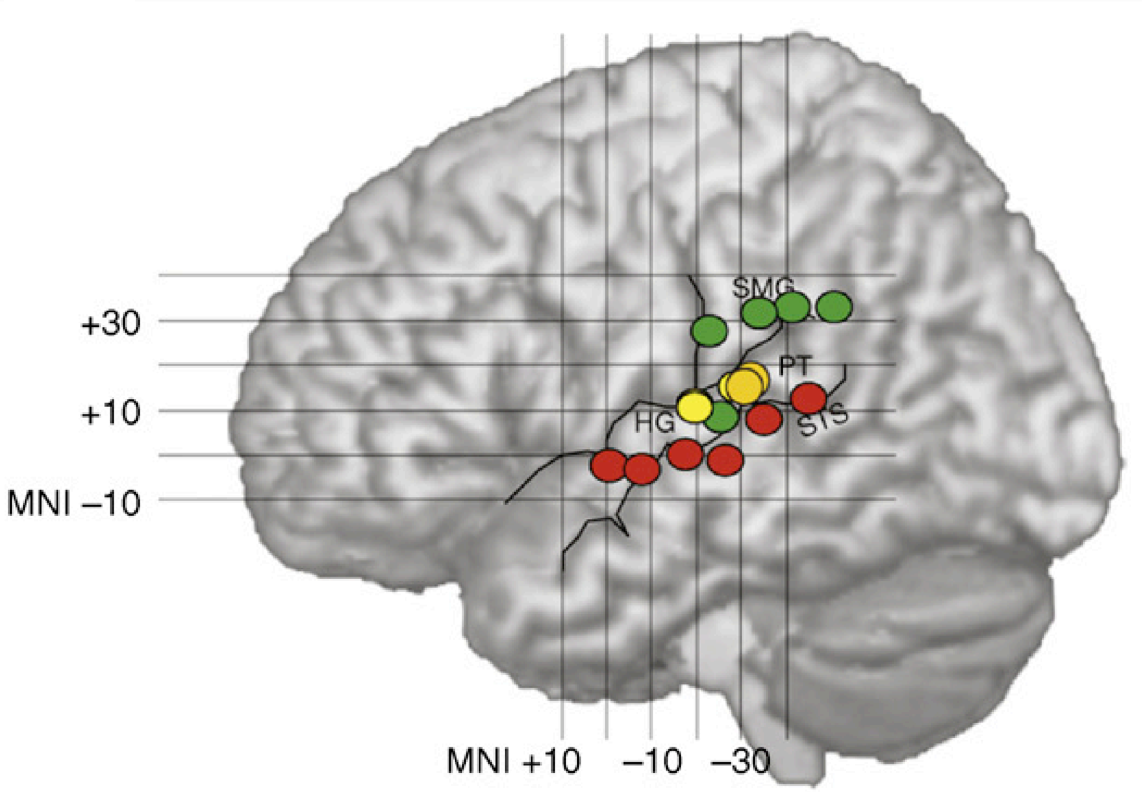
II. Perceptual learning

Adjusting pre-lexical categories

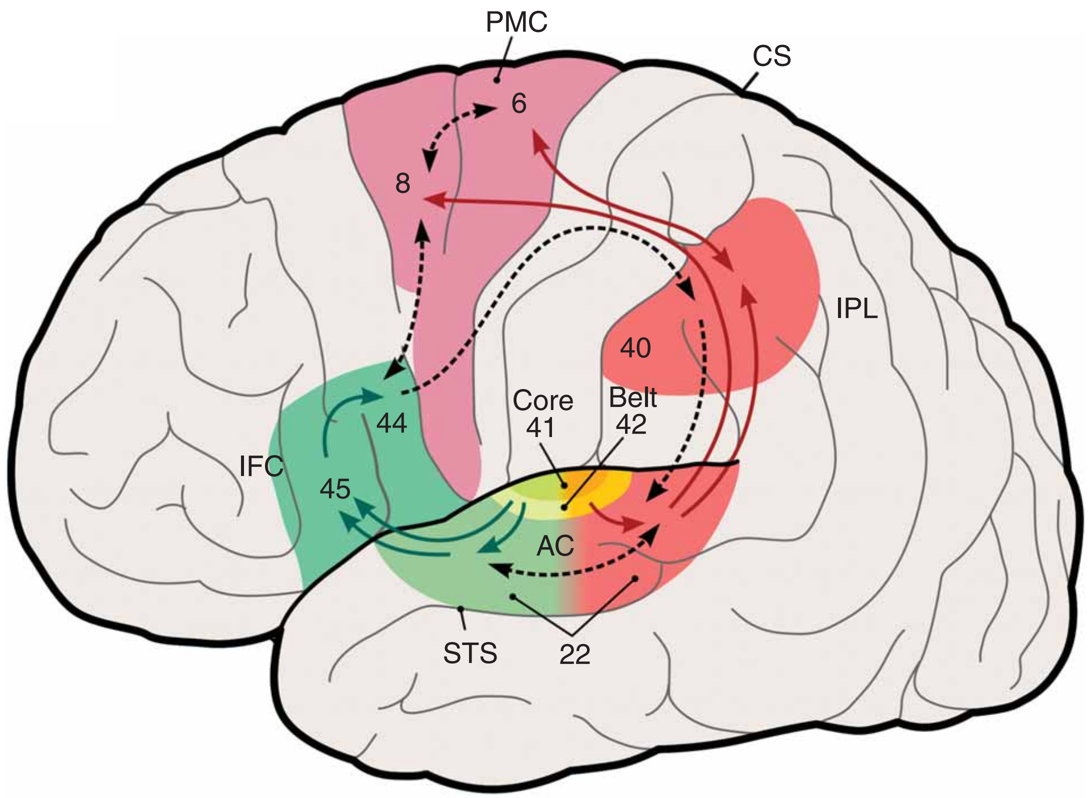
Adjusting to global signal degradation

III. Implications for models and applications

Pre-lexical vs higher-level processing



Obleser & Eisner 2009



Rauschecker & Scott 2009

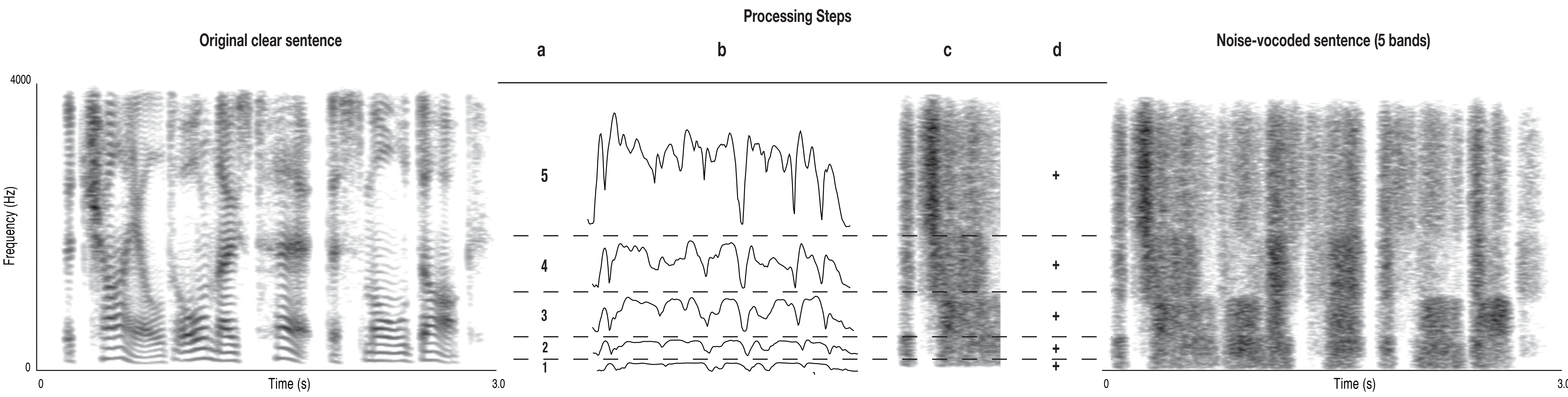
Increase in intelligibility

Improvement in speech comprehension after a period of exposure has been shown for various, ‘global’ sources of potential difficulty, including

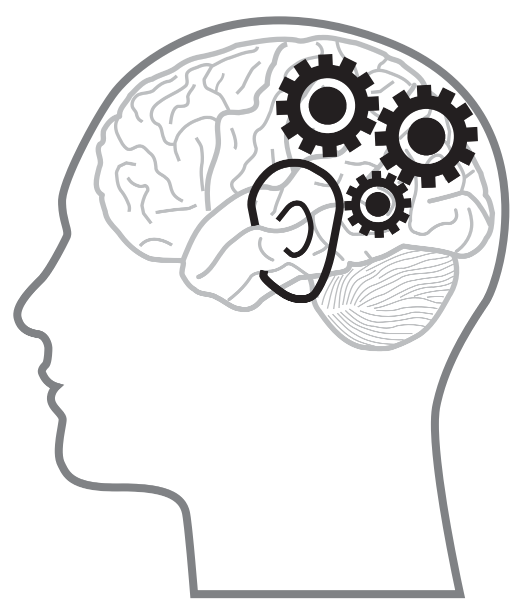
- accents — Clarke & Garrett 2004
- talker variation — Nygaard et al. 1998
- synthetic speech — Greenspan et al. 1988
- time-compressed speech — Dupoux & Green 1997
- noise-vocoded speech — Rosen et al. 1999

Noise-vocoded speech

Simulation of the stimulation received by users of a cochlear implant



Can individual differences in perceptual learning of this type of signal be predicted by **neural**, **cognitive**, or **auditory** measures?



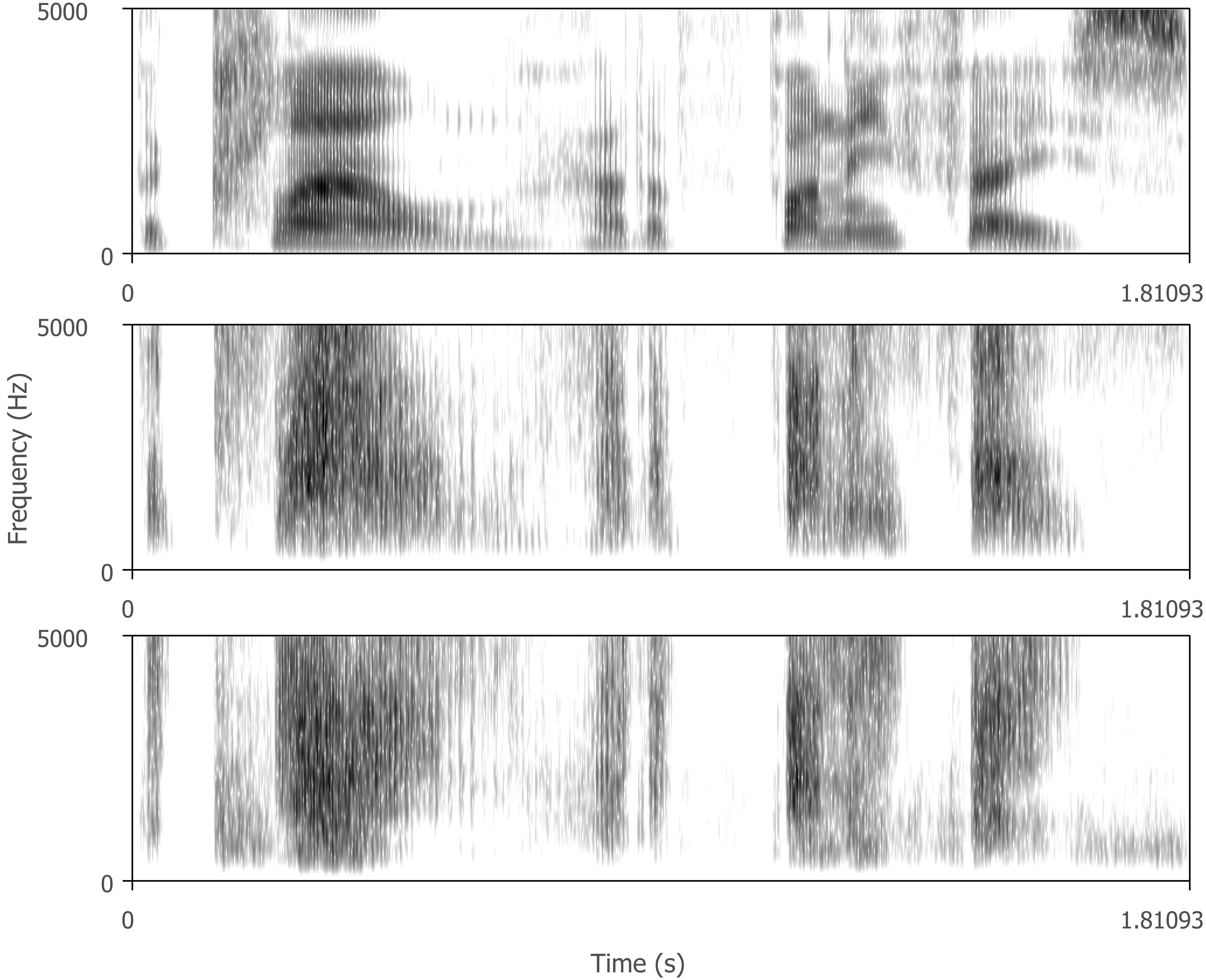
Noise-vocoded speech

Shannon et al. 1995; Rosen et al. 1999

Normal speech

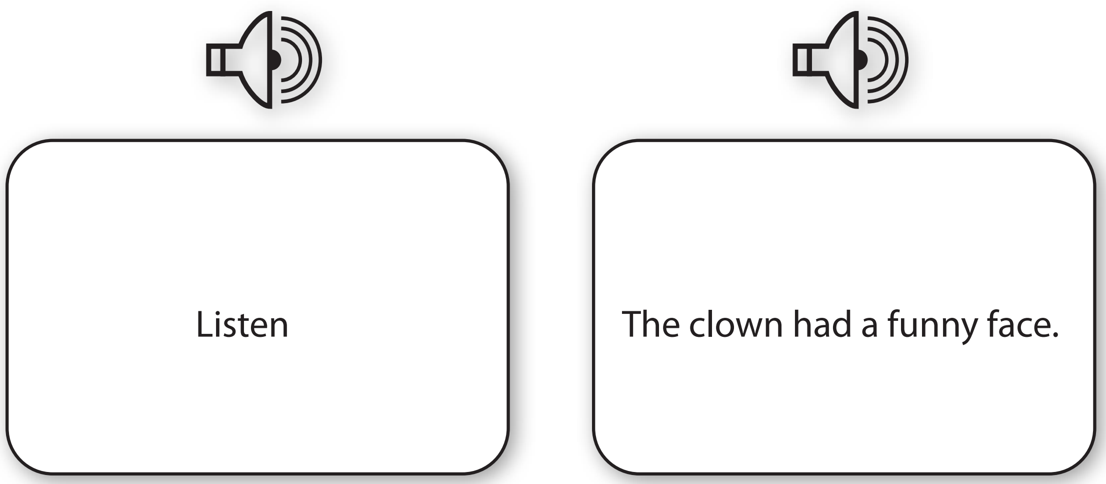
Noise-vocoded

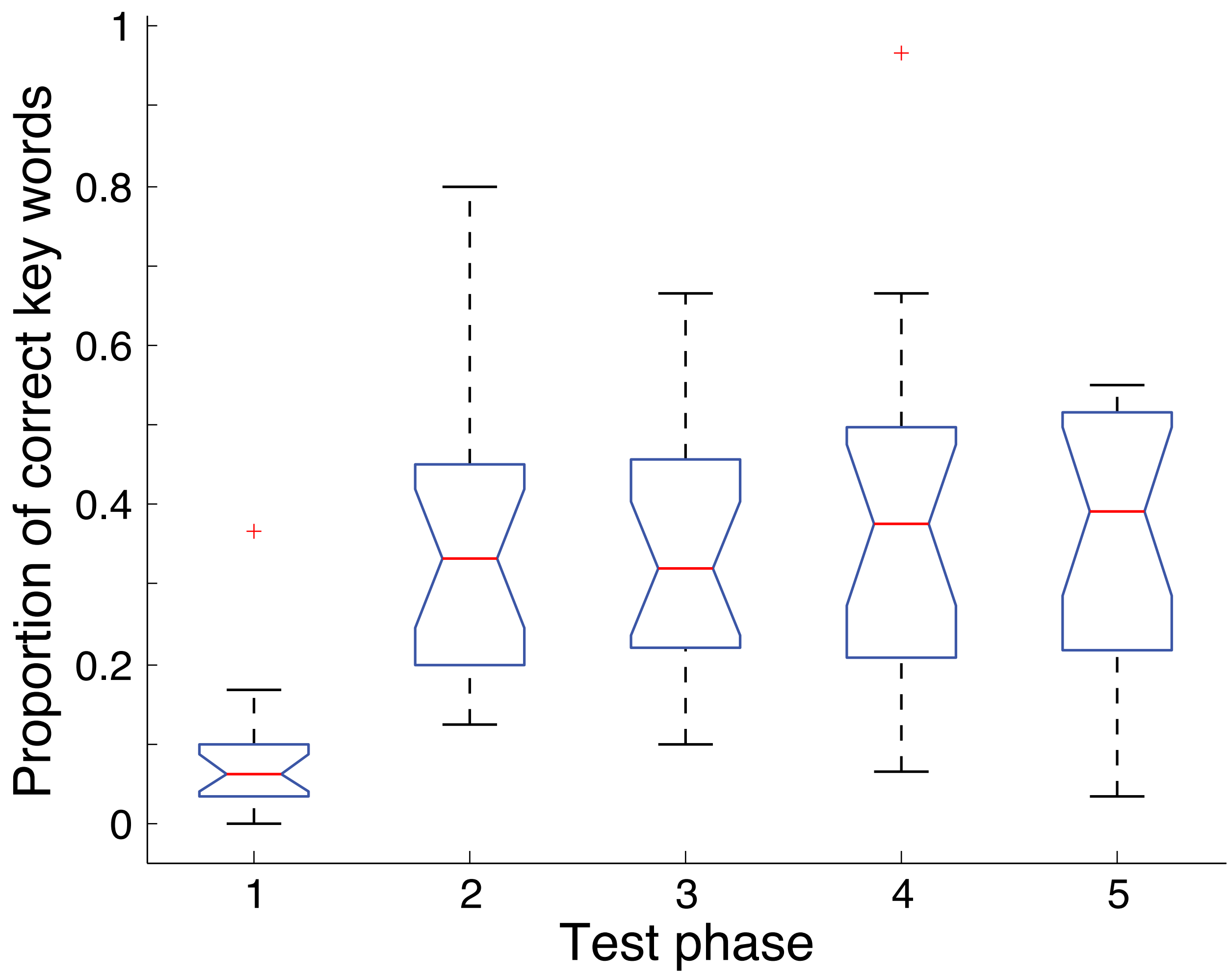
Noise-vocoded & spectrally inverted



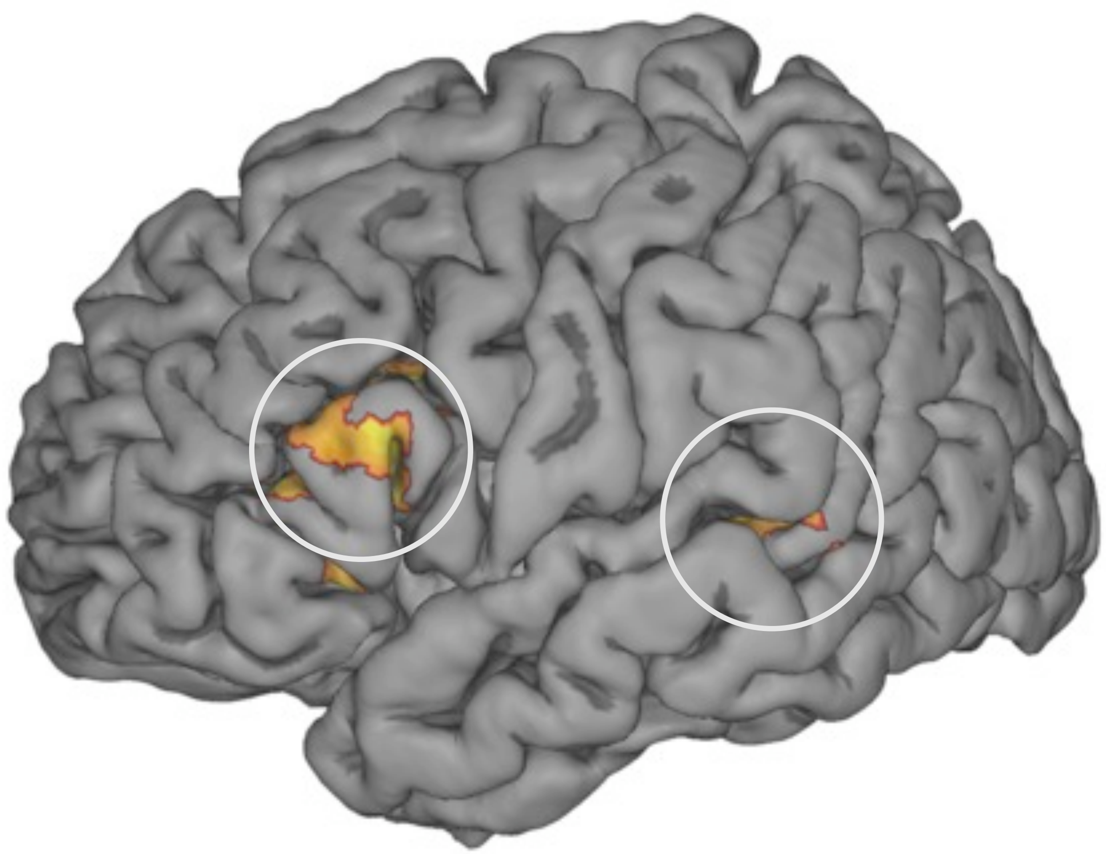
Training


Test block
Training block
time


Listen
The clown had a funny face.

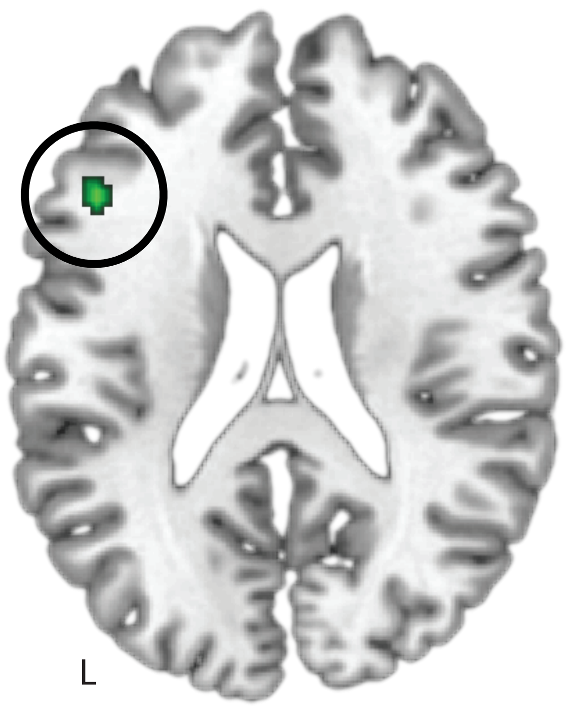

Proportion of correct key words
0
0.2
0.4
0.6
0.8
1
1
2
3
4
5
Test phase

Learnable vs. spectrally inverted sentences

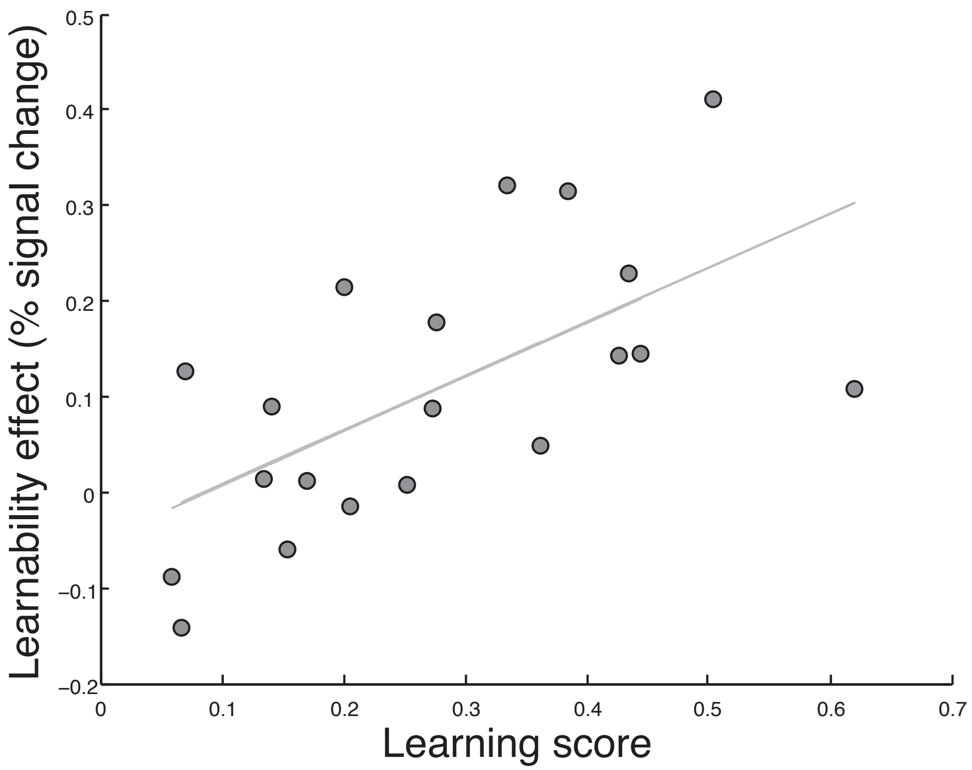


Eisner, McGettigan, Faulkner, Rosen, & Scott (2010)

Individual variation in performance



Inferior frontal gyrus (BA 45)



RFX; $p<.001$; $k=40$

Eisner et al. 2010

Learning to understand noise-vocoded speech

- Individual variation in overall performance and improvement, was associated with the left IFG, not auditory cortex.
- This pattern suggests that individual variation in learning to understand degraded speech does not arise from differences in acoustic-phonetic processing.
- Individual differences appear to be linked to the ability to use higher-level linguistic information to extract meaning from a noisy pre-lexical signal.

This talk

I. Models of speech perception

II. Perceptual learning

Adjusting pre-lexical categories

Adjusting to global signal degradation

III. Implications for models and applications

Applications

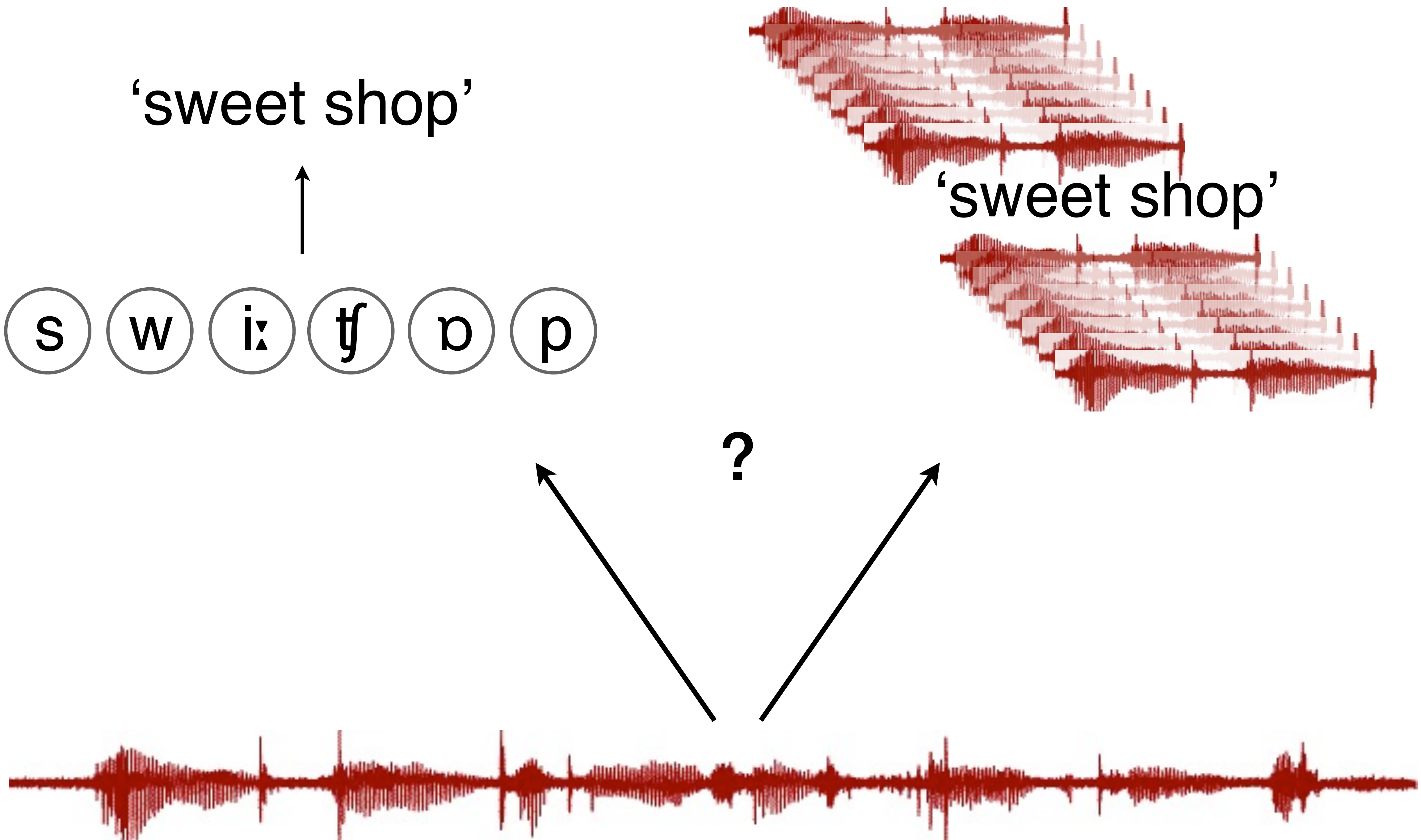
There is a great deal of plasticity in how we perceive speech.

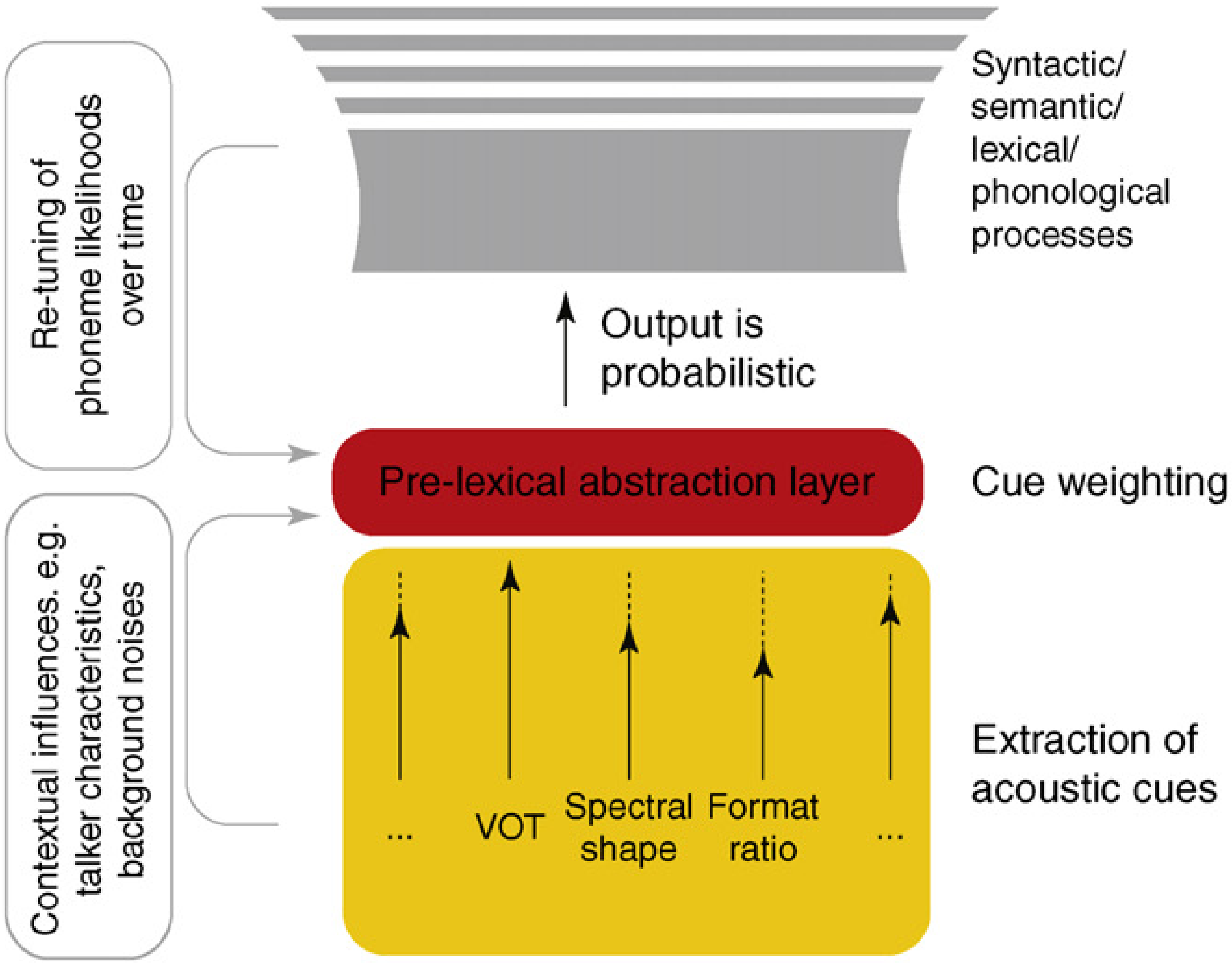
Relatively much is now known about how normal-hearing listeners cope with difficult speech input.

This could inform clinical applications in the domain of rehabilitation from hearing loss –

> for example, understanding causes of inter-individual variability can lead to more targeted training, both in terms of who needs specific treatment, and what training regimes are likely to maximise the potential for improvement.

Which is right?





Obleser & Eisner (2009) Trends in Cognitive Sciences.

Acknowledgements

Alissa Melinger
Department of Psychology
University of Dundee

Andrea Weber
Faculty of Humanities
University of Tübingen

Robin Iltzsche
Faculty of Humanities
Johann Wolfgang Goethe-University Frankfurt

Andy Faulkner
Department of Speech, Hearing and Phonetic Sciences
University College London

Stuart Rosen
Department of Speech, Hearing and Phonetic Sciences
University College London

Carolyn McGettigan
Institute of Cognitive Neuroscience
University College London

Sophie Scott
Institute of Cognitive Neuroscience
University College London

Jonas Obleser
Max Planck Institute for Human Cognitive and Brain Sciences

NWO
Netherlands Organisation for Scientific Research

wellcome trust

e-mail: f.eisner@ru.nl