

## Flexibility in speech perception

Anne Cutler

### Flexibility in speech perception

Why should speech perception be flexible?

ease of coping with novel talkers  
and novel listening conditions

Why should speech perception be inflexible?

advantage for the native language

### And, of course, language change

Language changes within a community,  
and individual talkers change along with it

The Queen's vowels changed from the 1950s to the 1980s perfectly tracking the changes in the language community. This production change must be driven by perception.

### Flexibility in speech perception: Outline

1. Perceptual flexibility for coping with novel talkers
2. Perceptual flexibility for coping with challenging listening conditions
3. L1 advantages

### Talker adaptation by perceptual learning

**1. LEARNING**  
(e.g. with lexical decision)

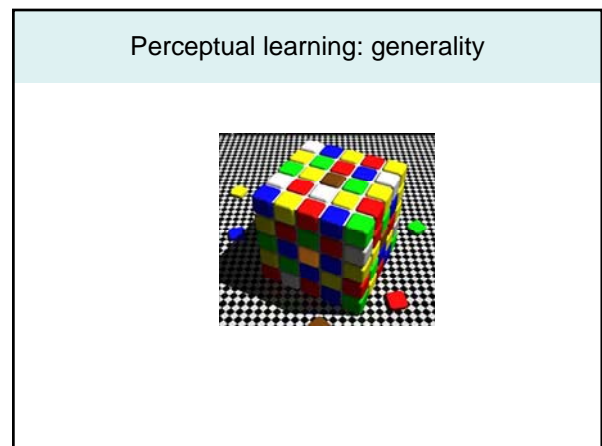
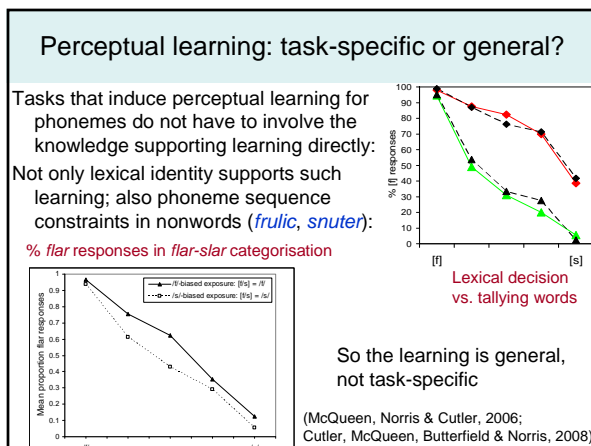
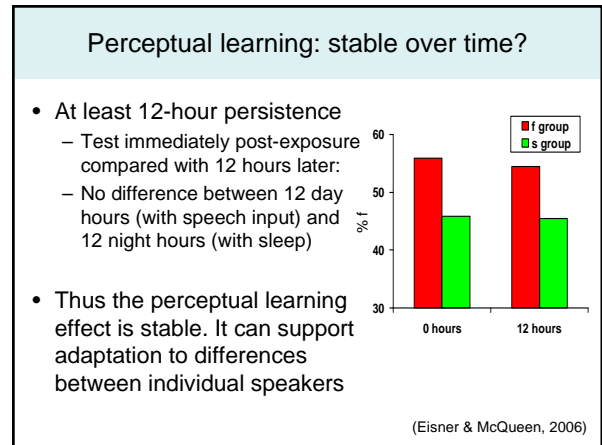
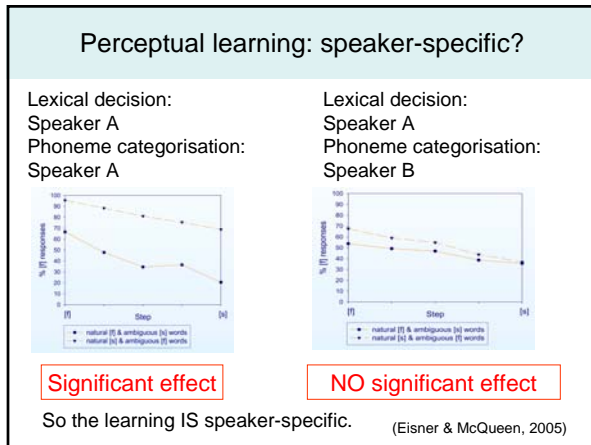
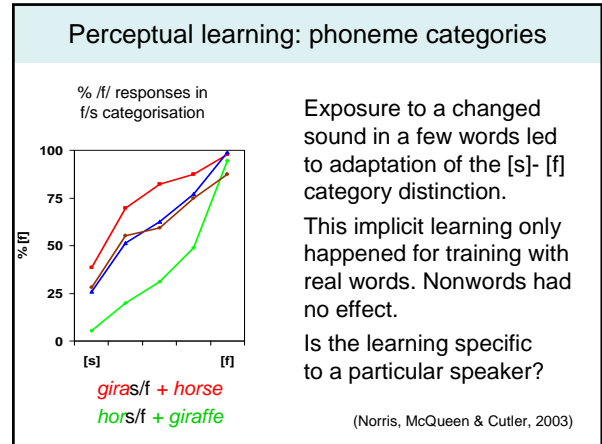
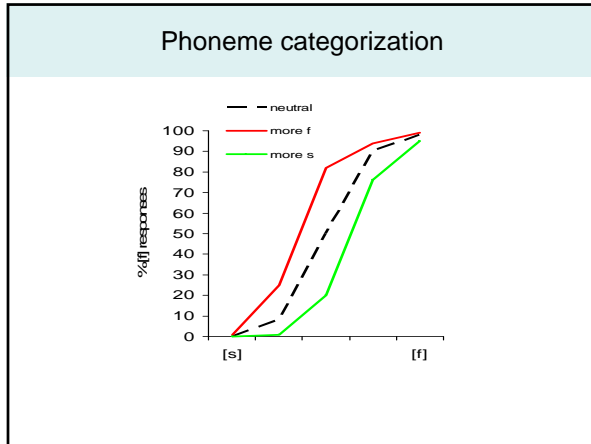
An altered sound [s/f] occurs instead of [f] in words like *gira[s/f]*, or instead of [s] in words like *hor[s/f]*

**2. TEST**  
(e.g. phoneme categorization)

Same [f/s] test for all



(Norris, McQueen & Cutler, 2003)






### Lexical decision



### Perceptual learning: generality

- **Two-part experiment** (from the subject's point of view, two separate experiments):
- (1) **Object recall**, including either:
 

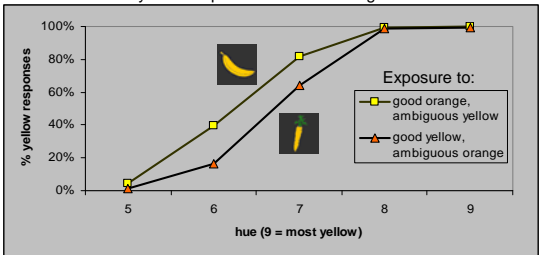

or:

- (2) **Colour categorisation**: orange or yellow
 

(Mitterer & De Ruiter, 2008)

### Perceptual learning: generality

% yellow responses in colour categorisation



Colour categories can be readily named, and they adapt to the input experienced, via reference to real-world knowledge.

(Mitterer & De Ruiter, 2008)

### Perceptual learning: generality

**Two-part experiment:**

(1) **Visual lexical decision**, including either:

WEIGH

or:

REIGH

(2) **Letter categorisation**:

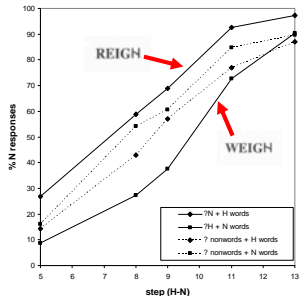
H or N? H H H H H

(Norris, Butterfield, McQueen & Cutler, 2006)

### Perceptual learning: generality

- In categorisation of colours and of letters, extraneous knowledge can be used to adjust category boundaries
- This is a powerful and general mechanism for perceptual learning
- Useful for perception of complex signals which arrive rapidly, overlap, and vary with context

% N responses in H/N categorisation




(Norris, Butterfield, McQueen & Cutler, 2006)

### Perceptual learning: does it generalise?


- Perceptual learning allows adaptation to new talkers, new dialects, and language change
  - Therefore it is (initially) speaker-specific
  - It is implicit, automatic and rapid
  - It is not dependent on a particular task
  - It is lasting across time
  - Crucially, it must generalise across words, i.e., not just hold for the words already heard

### Perceptual learning: generalisation


(1) Perceptual learning of an ambiguous sound as either /f/ or /s/:



(2) Interpret a word containing that sound

**KNIFE**  **NICE**

Priming: does *ni*[f/s] prime *knife* or *nice*?



(McQueen, Cutler & Norris, 2006)

### Perceptual learning: generalisation

<b>Spoken prime</b>	<b>Visual target</b>	Priming (= faster responses) if prime and target are the same. So: <b>f-group:</b> hears <i>n</i> [f/s] as <i>knife</i> ; → more priming for KNIFE <b>s-group:</b> hears <i>n</i> [f/s] as <i>nice</i> ; → more priming for NICE
<i>n</i> [f/s]	KNIFE	
<i>n</i> [f/s]	NICE	
<i>crop</i>	KNIFE	
<i>crop</i>	NICE	

i.e. the word consistent with the exposure in Part (1) should always receive the greatest facilitation

### Perceptual learning: generalisation

Priming effect: response facilitation against control

Words consistent with the Part 1 training were always more facilitated than inconsistent words, both in RTs and in errors.

So the perceptual learning generalised across words.

(McQueen, Cutler & Norris, 2006)

### Perceptual learning: generalisation

- The learning does generalise across words
- The effect is as strong as that of naturally spoken words (*nice*, *knife*)
- Learning can be acquired from novel sounds (e.g., a foreign phoneme replacing a native one), and then also generalises

Priming effect: response facilitation compared to control

(Sjerps & McQueen, 2010)

### Perceptual learning: positional generalisation?

Within speech categories, there can be positional specificity; a phoneme can take different forms in different positions. Does learning for an ambiguous sound in syllable-final position transfer to other positions in the word?

<b>spoken prime</b>	<b>visual target</b>
[f/s]eeling	FEELING
	CEILING

Yes, almost as strongly as for coda to coda transfer

(Jesse & McQueen, 2011)

### Early perceptual learning

**Part 1:** Picture verification:

**Part 2:** name continuum:

**Adults**

**12-year-olds**

**6-year-olds**

→ Even at age 6, listeners use perceptual learning to understand new talkers (and learn words from them...)

(McQueen, Tyler & Cutler, 2012)

### Perceptual learning across the lifespan

Even at age 70, with some age-related hearing deterioration, perceptual learning is still intact (even for fricatives).

Thus adaptation to new talkers is a constant component of speech processing across the whole lifespan.

(Scharenborg, Janse & Weber, 2012)

### Perceptual learning across speech categories

The results first found with fricatives replicate with stops (Kraljic & Samuel 2006), with liquids:

And indeed with tones, in Chinese:

(Mitterer, Chen & Zhou, 2011; Scharenborg, Mitterer & McQueen, 2011)

### Perceptual learning across speech categories

- Is all speech information equal?
  - equally susceptible to training?
  - equally informative in lexical access?
- Not always:
  - perceptual learning for vowels has been elusive
  - similar kinds of learning have been achieved with vowel manipulations (Maye, Aslin & Tanenhaus, 2008) but these are not implicit learning
- Relevant evidence: the word reconstruction task (hear a nonword; change ONE sound to reconstruct the real word)

### Reconstructing distorted words

**It is easier to modify vowels than consonants**

(Cutler, Sebastián-Gallés, Soler Vilageliu & van Ooijen, 2000)

### Perceptual learning across speech categories

- Listeners alter initial phonemic identity decisions more readily for vowels than for consonants
- Vowels vary more due to adjacent phonetic context; listeners have experience of this, and treat vowel information as less reliable
- In an implicit perceptual learning paradigm, a systematic vowel manipulation may get lost in the expected variability?
- In practice, consonants give more useful talker information than vowels, though in principle there is no difference in the way listeners process them

### Perceptual flexibility for coping with challenging listening conditions

- In native listening, flexibility at the phoneme category level supports talker adaptation and even language change
- There is also flexibility at the lexical level
- A major challenge in spoken-word recognition is rejecting words which are accidentally present in the input (word recognition contains *were*, *wreck*, *ignition*... but they should NOT be recognized!)
- A process of lexical competition allows the correct sequence of words to win out
- Several experimental tasks provide a view of the competition process

### Eyetracking

Participants hear speech input while looking at a display; where they look is monitored, e.g. by a head-borne camera:

The display typically contains referents that are temporarily compatible with the incoming speech.

		ham	kettle
		grapes	hammer

### Flexibility: Modulating lexical dynamics

Usually listeners are confident that speech sounds have been heard correctly. So words that begin in the same way are considered more seriously than words that begin differently (*candle* gets more competition from *candy* than from *handle*).

When there is noise around, this difference is greatly reduced (even for words not directly affected):

So, what words are considered is, in part, under listener control.

Condition	Onset match advantage
No noise	~17
Onset noises	~8
medial noises	~8

(McQueen & Huettig, 2012)

### Flexibility: Modulating lexical dynamics

Usually competing words overlap with the canonical form of a heard word. (So *beneden* gets more competition from *benadelen* than from *meneer*, even though in casual Dutch an initial [b] may become [m]).

This also alters, when there is reduced speech around (even for utterances that are not themselves reduced).

Again, the listener controls what lexical options are being considered.

Condition	Canonical form advantage
No reduction	~5
Some reductions	~1

(Brouwer, Mitterer & Huettig, 2012)

### Phonetic and lexical flexibility in listening

- (Native) listeners can adjust
  - the boundaries of their phonetic categories
  - the competitor population as they recognize words
- This flexibility in adjusting the parameters of the processes making up spoken-language recognition is arguably responsible for multiple known cases of L1 advantage:
  - in talker identification
  - in listening under noisy conditions
  - in adaptation to new accents

### Talker identification

- Long known: Identifying talkers is easier in L1.
- E.g.: the same set of English-German bilinguals are distinguished better by English-speakers if speaking English, but by German-speakers in German:
- Is this due to how well the speech is understood?

Listener Group	English speech	German speech
English listeners	~40	~10
German listeners	~35	~55

(Thompson, 1987; Goggin, Thompson, Strube & Simental, 1991; Schiller & Köster, 1996; Schiller, Köster & Duckworth, 1997)


### Beginnings of talker perception

- Talker perception starts early
- – preference for mother's voice at birth
- What about new talkers?
- When do we become able to tell the difference between talkers and notice a talker switch?
- Discrimination can be tested in babies with a habituation/test paradigm

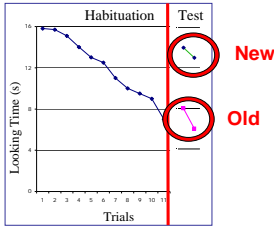
### Testing early discrimination

- 1. HABITUATION**
  - Repeating stimulus
- 2. TEST**
  - Stimulus changes;
    - is the change noticed?

### Testing early discrimination



Dependent measure: looking time

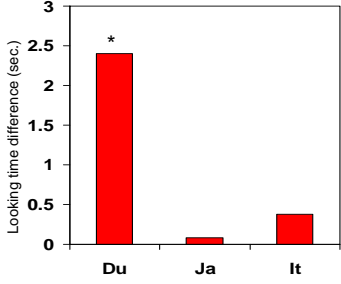


Infant-controlled Habituation Paradigm

Can infants discriminate between unfamiliar talkers uttering sentences? (e.g., *Artists are attracted to life in the capital...*)

### Discriminating between talkers at 7 months?

Discrimination = looking time to Test trials longer than looking time to last two Habituation trials

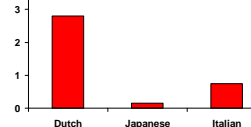


Only in the native language!

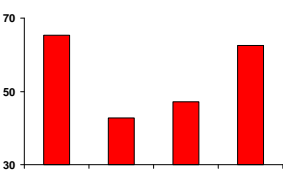
(Johnson, Westrek, Nazzi & Cutler, 2011)

### Talker identification in L1 and a foreign language

Dutch-acquiring 7-month-olds discriminate talkers in Dutch but not in other languages:



English-speakers identify talkers better in English speech than in unfamiliar languages:



A familiar (albeit not comprehensible) language is almost as easy: i.e. the phonology suffices

(Johnson, Westrek, Nazzi & Cutler, 2011)

### Talker discrimination, identification, adaptation

- Infants can discriminate between talkers and notice a talker switch (in the native language)
- Identification requires greater memory skills and the ability to form abstract voice categories
- Some phonemes more useful than others (how fast voices are learned depends on what is said!)
- There are voice-selective areas in the brain, closely tied to language processing areas
- (See PhD thesis by Attila Andics, 2013, for much more!)

### 2. Listening in noise

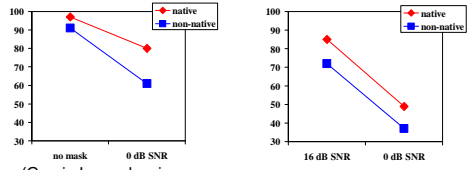
- The clearest way to see a difference between non-native and native listening!
- 2 attempts to view this strictly at phoneme level:

- American English (all 645 possible CV or VC syllables)
- Multi-speaker babble masking
- Dutch (highly proficient) non-native listeners, American English native listeners

- American English: 16 consonants in 160 aCa tokens
- Multi-speaker babble masking
- Spanish (moderately proficient) non-native listeners, British English native listeners

(Cutler, Weber, Smits & Cooper, 2004; Garcia Lecumberri & Cooke, 2006)

### First language advantage: Listening in noise



(Garcia Lecumberri & Cooke, 2006)      (Cutler, Weber, Smits & Cooper, 2004)

- Possible explanations: (a) Dutch non-native listeners are too good (but then why aren't they as good as native listeners?)
- (b) Even constant timing and a constant vowel context offers a predictability advantage (that native listeners can use)



### First language advantage: Listening in noise

- Test: give Dutch listeners the materials used by GLC
- If the parallel native-Dutch performance is due to the proficiency of Dutch non-native listeners, results will be parallel again
- If the parallel native-Dutch performance is due to absence of predictability cues (that natives can use better), results will now NOT be parallel

The crucial difference is that L1 listeners have the resources to recover from the effects of noise.

(Cutler, Garcia Lecumberri & Cooke, 2008)

### First language advantage: Listening in noise

Understanding speech in noise is much harder in a second language than in the native language.

Again, this effect can be largely accounted for phonologically

(Cutler, Weber, Smits & Cooper, 2004; Cutler, Garcia Lecumberri & Cooke, 2008; Garcia Lecumberri, Cooke & Cutler, 2010)

### Native and non-native listening in noise: Vowel and consonant identification

Highly significant positive correlation ( $r = .91$ ) between percent correct recognition per phoneme by native (vertical axis) and non-native listeners (horizontal axis)

### 3. Adapting to other pronunciations

- Flexibility at the phonetic category level should allow adjustment to different category realization in another L1 dialect – and it does:
- Flexibility at the level of lexical dynamics should mean that the competitor population can be adjusted to suit the dialect being heard (not yet directly tested! Some supporting evidence)

(Dahan, Drucker & Scarborough, 2008; Trude & Brown-Schmidt, 2012) (Cutler, Smits & Cooper, 2005)

### Adapting to other pronunciations

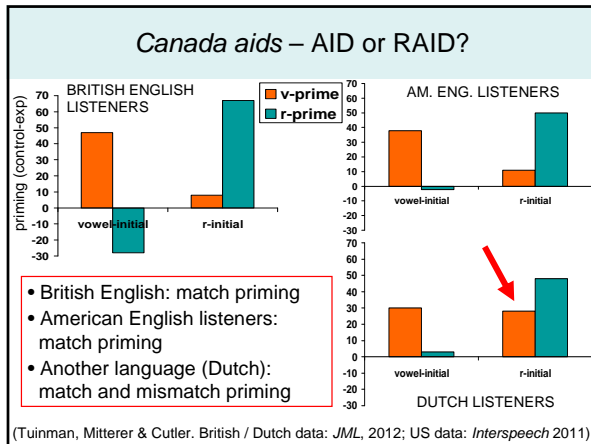
- Unfamiliar features in a dialect of the L1 (with no equivalent in the native variety)?
- Example 1: [r]-intrusion in British English (e.g., *saw* [r] *a film*)
- Can listeners with another English dialect tell real from intrusive [r]? (The crucial clue is duration... Does spelling distract? (*saw* [r] *ice* vs. *more* [r] *ice*))
- Does intrusive [r] activate unintended words? Is *Canada aids* heard as *Canada raids*? (i.e., does a sentence containing *Canada aids* prime recognition of AID or of RAID?)

### Real vs. intrusive [r] – *ice* or *rice*?

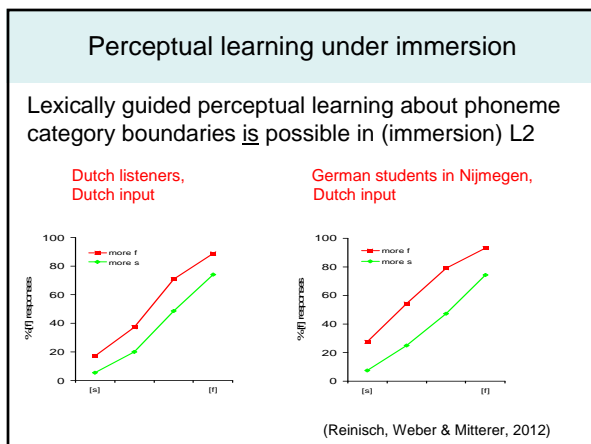
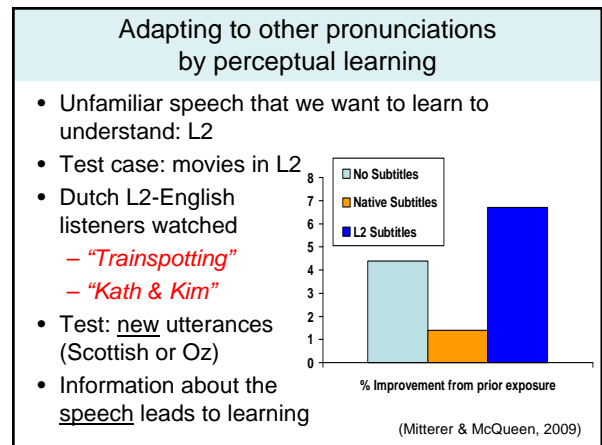
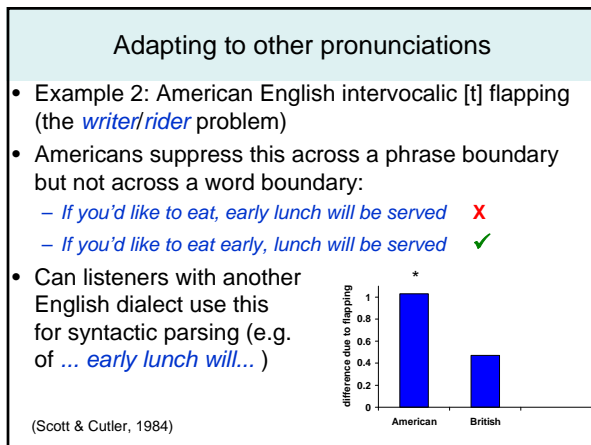
- British English listeners attend to [r] duration only
- Listeners from another language (Dutch) rely mainly on spelling
- Listeners from another variety of English (AmE) are in between

(Tuijnman, Mitterer & Cutler, 2011)





- ### Adapting to other pronunciations
- Unfamiliar features in a dialect of the L1 do not lead L1 listeners to mistakenly recognize spurious words
  - Even though these dialectal features are not dealt with fully efficiently (at the phoneme level)
  - Or at other levels (e.g., phrase boundary detection)
  - Perceptual learning can improve adaptation to such dialectal features
  - Perceptual learning is a continuum (e.g., from one-off talker adjustment to lasting language change)
  - Is immersion continuous perceptual learning?



- ### Using perceptual learning mechanisms in L2
- Dutch listeners to English have great trouble distinguishing the vowels in *cattle* vs. *kettle*.
  - But their phonological representations in the lexicon are distinct – so they have used other information (e.g., spelling) to keep them apart.
  - Training such listeners to label nonsense items with nonsense “English” names such as *tendik*, *tandis* produces homophonous first-syllable representations is they only hear the names, but distinct representations if they can also read them.
- (Weber & Cutler, 2004; Escudero, Hayes-Harb & Mitterer, 2008)

### Adapting to other pronunciations

- Example 2: American English intervocalic [t] flapping (the *writer/rider* problem)
- Americans suppress this across a phrase boundary but not across a word boundary:
  - If you'd like to eat, *early lunch will be served* ❌
  - If you'd like to eat *early*, *lunch will be served* ✅
- Can listeners with another English dialect use this for syntactic parsing (e.g. of ... *early lunch will...* )

(Scott & Cutler, 1984)

### Flexibility in speech perception

- Flexibility is multi-faceted
- We adapt to novel talkers and listening conditions
- But rigidly stay within the confines of the L1
- The L1 advantages are in part due to greater flexibility (adaptation, recovery) in L1 than in L2
- Yet we can use the same adaptation techniques to adapt to new L2 talkers (as well as we do in L1??)
- And maybe the effects of immersion resemble the way language change takes place
- Modulation of L2 lexical dynamics? Not yet known...
- Room for a ton of research here!!!!!!!!!!!!!!!

### Native Listening: an advertisement

How listening to spoken language is so efficient – it's because listening is tailored to the native language

MIT Press, 2012

### Modeling perceptual learning in MINERVA2

- MINERVA2 – word recognition episodes produce independent traces; inputs cause traces to echo
- Traces modeled as vectors of 400 binary elements: 200 *name* elements for category identity, and 200 *form* elements for stimulus properties
- 500-word lexicon; 40 “s-words” (*horse, nice* etc.), 40 “f-words” (*giraffe, knife* etc.); 20 minimal pairs
- Training: 20 ambiguous forms for *horse*-words, plus 20 unambiguous *giraffe*-word forms (or vice versa)
- Test: forms ambiguous between *knife* and *nice*
- Output echo content more similar to *knife* or *nice*?

(Cutler, Eisner, McQueen, & Norris, 2010)

### Effects of Training: Humans vs. Model (Minimal pair interpretation consistent with training?)

Human listeners learn effortlessly, from just a few exemplars. MINERVA2 predicts the reverse effect. For listeners, ambiguous sounds are phonemes, but for the model they are nothing.

	consistent	inconsistent
Human (RTs)	~80%	~20%
Human (Item Errors)	~20%	~80%
Model	~50%	~50%
Model – 10x training	~80%	~20%

### Voice-selectivity in the brain

- Neural representations involve voice-selective areas, voice-specific norms, and flexibility
- The flexibility allows rapid adaptation to new voices (for understanding novel talkers)

(Andics, Turenout & McQueen, *ICPhS* 2007; Andics, McQueen, Petersson, Gál, Rudas & Vidnyánszky, *NeuroImage* 2010; Cutler, Andics & Fang, *ICPhS* 2011; Andics, 2013)