

MIT OpenCourseWare
<http://ocw.mit.edu>

24.963 Linguistic Phonetics
Fall 2005

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

24.963

Linguistic Phonetics

Speech Perception and Lexical
Access

- Final papers are due on Mon Dec 19th.
- Any missing assignments are due immediately.

A role for phonetic/phonological grammar in speech perception: parsing contextual effects

- Gaskell and Marslen-Wilson (1996) suggest that listeners use knowledge of assimilation processes to ‘undo’ assimilation in the process of lexical access.
 - $n \rightarrow m / _ [\text{labial}]$
 - $[\text{limbeik}\eta] \rightarrow \text{linbeik}\eta/\text{limbeik}\eta \rightarrow \text{‘lean bacon’}$
- Alternative accounts of lexical access in the face of significant (near-neutralizing) variation in word form:
 - Permissive matching between stimulus and lexicon.
 - Underspecified lexical representations (Lahiri and Marslen-Wilson).
 - Feature parsing (Gow)

Permissive matching in lexical access

- [lim] is close enough to [lin] to activate ‘lean’, resulting in access since this is the best match in the lexicon.
- [leɪk(kɔl)] is close enough to [leit] to activate ‘late’, ultimately resulting in access since this makes more sense than competitor ‘lake’.
- These mechanisms cannot account for evidence that activation is greater in environments where word form variation is the result of grammatical variation:
 - G&M-W ‘96: [wikɪb pɹæŋk] primes ‘wicked’ more than [wikɪb geɪm].
 - If [wikɪb] activates ‘wicked’ because it is ‘close enough’, then this effect should be similar in both contexts.

Underspecified lexical representations

- Lahiri and Marslen-Wilson (1991, 1992) proposed that lexical representations lack features specifications in certain contexts.
- In English [coronal] is not specified.
- So any place specification in the input will match the lexical representation of a coronal - [lim], [lin], [liŋ] activate ‘lean’ equally.
- This is similar to permissive matching (matching is strict, but some lexical representations are ‘permissive’), so is not sufficient to account for cases in which activation by a non-canonical form is stronger in assimilatory contexts.
- NB Lahiri et al (2002) did find that in German *Bah*[m] primes visual *Zug* ‘train’ as much as canonical *Bahn* ‘railway’

Feature parsing

- Gow (2001, 2002) shows that English coronal assimilation is not complete - the result retains some cues to the presence of an alveolar (and some articulatory remnants of the alveolar - Nolan 1992).

Feature parsing:

- [liⁿm...] contains cues to both [coronal] and [labial].
- On hearing following [b], [labial] is attributed to [b], and [coronal] is assigned to the nasal, resulting in [lin].
- If there is no following labial (limgeim), [labial] and [coronal] cannot both be accommodated - if evidence for [labial] is stronger, perceive [lim].
- Feature parsing is argued to be a general perceptual mechanism - not based on knowledge of coronal assimilation.
- Not applicable to total assimilation as in Gaskell and Marslen-Wilson's stimuli.
- Not applicable to non-assimilatory phenomena (e.g. deletion/extreme reduction).

Post-lexical vs. Pre-lexical processes

- Do not confuse with lexical and post-lexical phonology!
 - Pre-lexical process is one that occurs before matching to lexical representations.
 - Post-lexical process is one that occurs after lexical access (or at least after candidates for lexical access have been identified).
- Gaskell and Marslen-Wilson propose that listeners use knowledge of phonological processes pre-lexically - i.e. parsing the input to derive representations for lexical matching.
- Could phonological rules be applied to choose between word candidates after lexical access?
 - [wikibgeim] initially activates 'wicked', but suppressed when following context is phonologically inappropriate.

Processing contextual variation

- Multiple mechanisms could be at work in processing substantial contextual modifications (like assimilation).
- E.g. [lim] activates [lin] to some extent in non-assimilatory contexts, but less than in assimilatory context.
- G&M-W (1998) found that listeners ‘recovered’ [t] by undoing assimilation in non-words in a phoneme-monitoring task
- suggests pre-lexical parsing
- but they recovered /t/ more often in real words, suggesting there is also a lexically-based mechanism involved.

Darcy et al (2004)

Darcy, Ramus, Christophe, Kinzler & Dupoux (2004), Ms.

- Used cross-language comparison to test whether compensation for assimilation depends on knowledge of the assimilation process, or is based on general perceptual mechanisms (cf. feature parsing).
- Regressive voicing assimilation in French, but not English.
robe [ʁɔb] robe sal [ʁɔp sal]
- Coronal assimilation in English, but not French.
moine [mwan] moine bavard *[mwambavaʁ]
- If knowledge of phonology is involved in compensation for assimilation, then opposite patterns should be observed in the two languages.

Darcy et al (2004)

- Two types of target: canonical word, changed (non-word).
- Three conditions:
 - viable change - change is assimilation to context (not necessarily viable in French) [kɔp sal], [lym pal]
 - unviable change - non-assimilatory change [kɔp nwak], [lym pal]
 - no change - canonical form in a non-assimilatory context (sonorant, k for voicing; coronal, k, f for place (?). [kɔp kuz], [lyn zɔn]
- Task: word spotting (target presented aurally, no response if word is mispronounced).

Darcy et al (2004)

- Minimal compensation for place assimilation (illegal in French)
- Substantial compensation for voicing assimilation (legal in French).

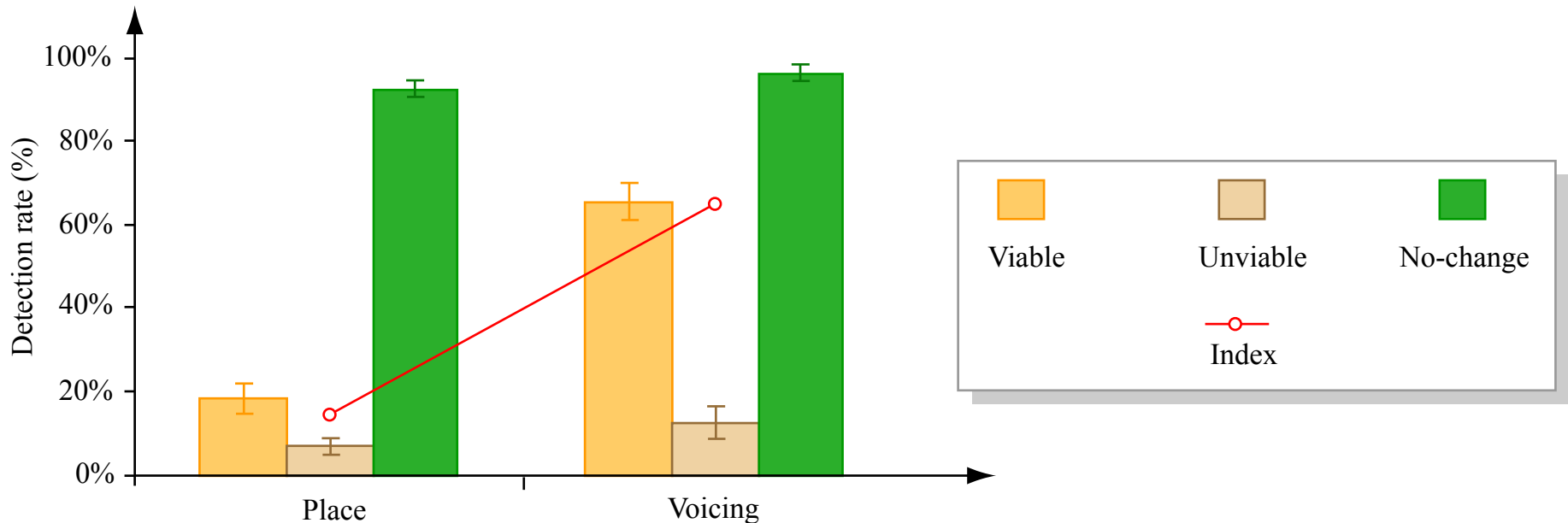
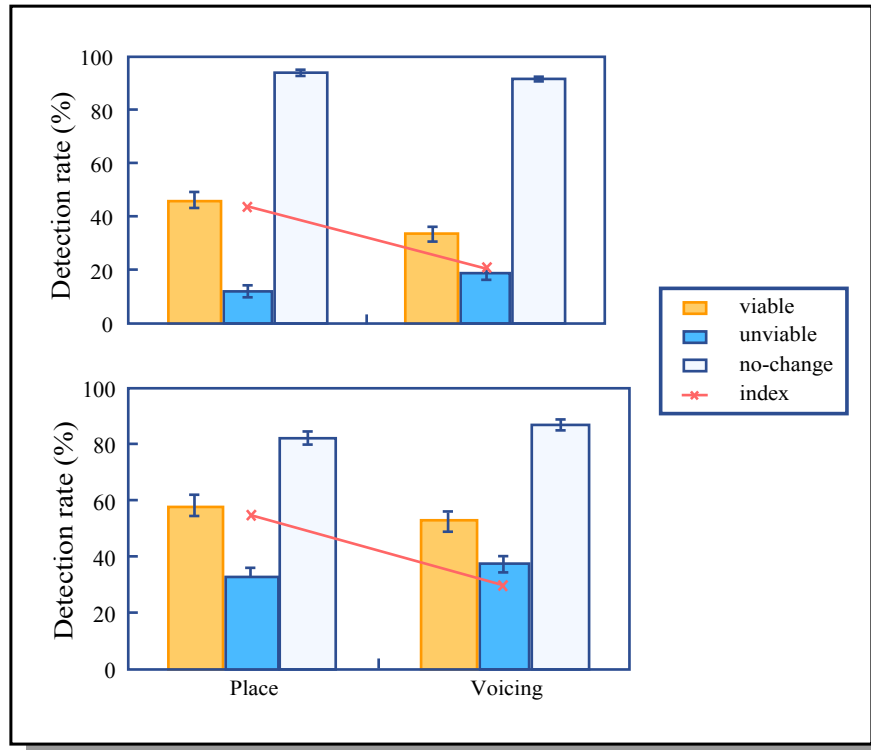


Image by MIT OpenCourseWare. Adapted from Darcy, I., F. Ramus, A. Christophe, K. Kinzler, and E. Dupoux, E. "Phonological Knowledge in Compensation for Native and Non-native Assimilation." In *Variation and Gradience in Phonetics and Phonology*. Edited by F. Kinsler, C. Fery, and R. van de Vijver. New York, NY: Mouton De Gruyter, forthcoming.

Darcy et al (2004)

- Two further experiments of the same kind:



American
English (subjects
and stimuli)

English learners
of French, French
stimuli

Image by MIT OpenCourseWare. Adapted from Darcy, I., F. Ramus, A. Christophe, K. Kinzler, and E. Dupoux, E. "Phonological Knowledge in Compensation for Native and Non-native Assimilation." In *Variation and Gradience in Phonetics and Phonology*. Edited by F. Kinsler, C. Fery, and R. van de Vijver. New York, NY: Mouton De Gruyter, forthcoming.

Darcy et al (2004)

- The extent of compensation for assimilation is dependent on experience with the assimilation process.
- There is some compensation for improper assimilation.
 - Darcy et al suggest that universal compensatory mechanisms are involved.
 - Some degree of devoicing assimilation probably occurs in English.
 - ‘Viable’ contexts for assimilation always involve obstruents - worse environment for the perception of place and obstruent voicing than pre-sonorant (included in ‘unviable’ contexts).
- It has been shown that many kinds of assimilation are more likely in environments where the relevant contrast is more difficult to discriminate.
- Implies a confound between attested patterns of assimilation and discriminability - makes it difficult to test claims about universal mechanisms of compensation for assimilation.
- Perhaps other processes would not suffer this problem?

Mitterer, Csépe & Blomert (2003)

- Test effects of experience on ability to compensate for assimilation, using native and non-native listeners.
- Test for pre-lexical compensation using non-words.
- Hungarian liquid assimilation: /l/ → [r] / _ r
 - bɔlro:l / bɔlrro:l ‘from the left’
 - bɔlna:l / *bɔrna:l ‘at the left’
- Constructed bɔl-bɔr (word-nonword) and zɔl-zɔr (both non-words) continua through resynthesis.
- Spliced onto ro:l or na:l
- Task: identification as bɔl-bɔr / zɔl-zɔr (orthographic labels) (better than phoneme monitoring?)
- Subjects: Hungarian and Dutch speakers.

Results

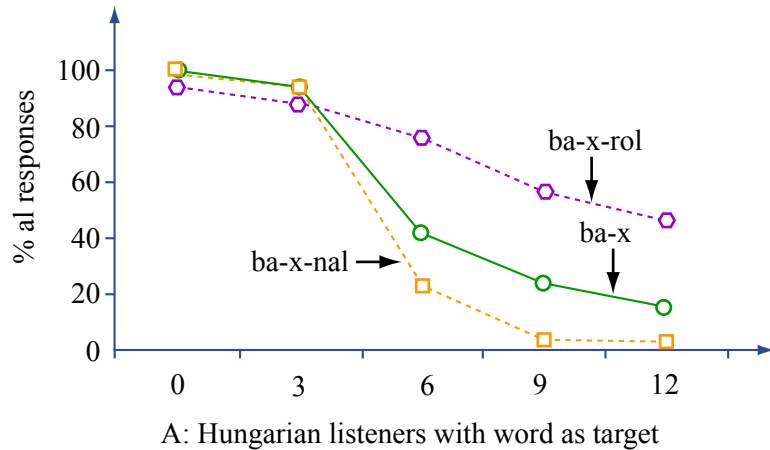


Image by MIT OpenCourseWare. Adapted from Mitterer, H., V. Csepe, and L. Blomert. "The role of Perceptual Integration in the Perception of Assimilation Word Forms." *Quarterly Journal of Experimental Psychology* 59 (2006): 1395-1424.

- More compensation in viable contexts for words and non-words - pre-lexical effect.
- But non-word does pattern differently from word continuum before 'nal'

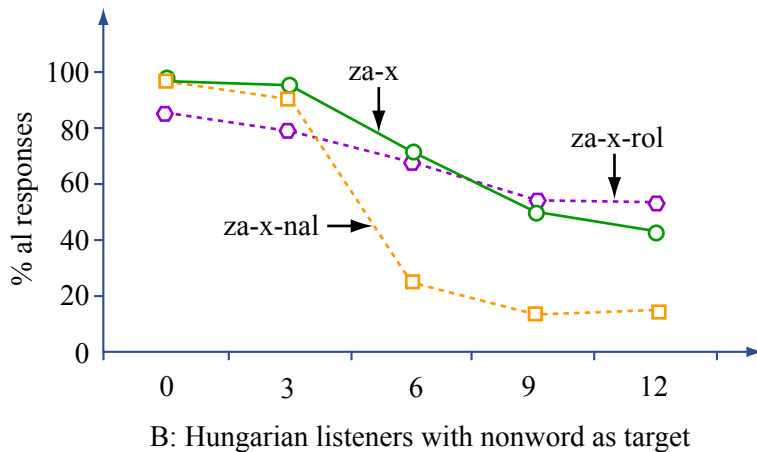
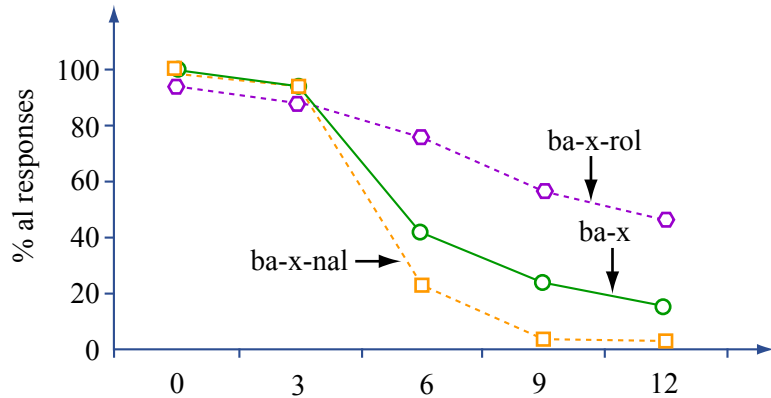


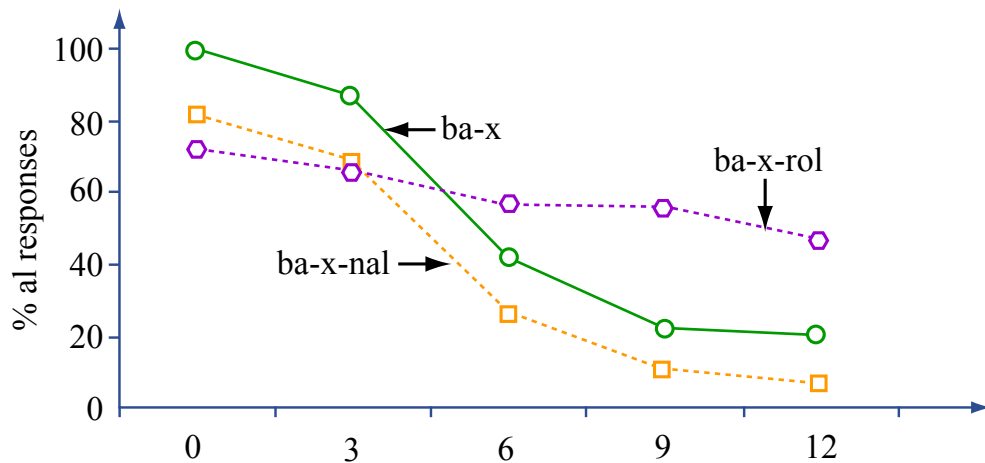
Image by MIT OpenCourseWare. Adapted from Mitterer, H., V. Csepe, and L. Blomert. "The Role of Perceptual Integration in the Perception of Assimilation Word Forms." *Quarterly Journal of Experimental Psychology* 59 (2006): 1395-1424.

Results



A: Hungarian listeners with word as target

Image by MIT OpenCourseWare. Adapted from Mitterer, H., V. Csepe, and L. Blomert. "The Role of Perceptual Integration in the Perception of Assimilation Word Forms." *Quarterly Journal of Experimental Psychology* 59 (2006): 1395-1424.



C: Dutch listeners with Hungarian word as target

Image by MIT OpenCourseWare. Adapted from Mitterer, H., V. Csepe, and L. Blomert. "The Role of Perceptual Integration in the Perception of Assimilation Word Forms." *Quarterly Journal of Experimental Psychology* 59 (2006): 1395-1424.

- Dutch listeners perform differently from Hungarian listeners.
- Mitterer et al conclude that Dutch listeners show partial compensation for assimilation.
- It looks more like simple confusion in the 'rol' context.

Summary

- Good evidence that listeners use their knowledge of assimilation processes to process assimilated forms, not just general purpose perceptual processes.
- Some evidence that this knowledge is used in pre-lexical parsing of stimuli.
- Should look at processes other than assimilation, e.g. deletion, lenition.

Word segmentation

- There are generally no spaces between words in speech.
- So lexical access involves segmentation of the speech stream as well as identification of words.
- We have been considering the problem of performing lexical access based on an acoustic word form, but how do we know which chunks of the acoustic stream to attempt lexical access with?

Word segmentation

Two basic strategies:

- Use word boundary cues to identify word onset before beginning lexical access.
- Lexical - segmentation as a side-effect of recognition
 - Sequential: after a word has been recognized, initiate lexical access again. E.g. Marslen-Wilson's Cohort model.
 - Problem: onset-embedded words, e.g. *hamster-ham*.
 - Parallel: Attempt lexical access starting at frequent intervals (e.g. every segment), entertaining multiple segmentation hypotheses. Select best word sequence. E.g. McClelland and Elman's Trace model.

Word boundary cues

- Classic study: Nakatani and Dukes (1977).
- Studied ambiguous phoneme sequences

Juncture consonant	Parent phrases ^a			
	Juncture before consonant	% correct juncture perception	Juncture after consonant	% correct juncture perception
/d/	stay dill	53.3%	stayed ill	93.3%
/t/	play taught	100%	plate ought	100%
/q/	you thread	46.7%	youth read	60.0%
/z/	buy zinc	33.3%	buys ink	100%
/g/	lawn chair	100%	launch air	93.3%
/n/	no notion	53.3%	known ocean	93.3%
/l/	we loan	100%	we'll own	93.3%
/r/	two ran	100%	tour an	100%

Image by MIT OpenCourseWare. Adapted from Nakatani, L. H., and K. D. Dukes. "Locus of Segmental Cues to Word Juncture." *Journal of Acoustical Society of America* 62 (1977): 714-719.

Word boundary cues

Image removed due to copyright restrictions.

See Nakatani, L. H., and K. D. Dukes. "Locus of Segmental Cues to Word Juncture." *Journal of Acoustical Society of America* 62 (1977): 714-719. □□□□

- Divided each sequence into 4 parts.
 - Word 1 up to transition into medial C
 - Transition to middle of C
 - Transition from middle of C
 - Rest of word 2.
- Cross-spliced the pieces of minimal pairs in every possible combination (16).
- Subjects identified word sequence, four alternative forced choice, e.g.
 - no ocean
 - no notion
 - known ocean
 - known notion

Word boundary cues

- Segmentation did not depend on initial or final portions, except in the case of final portion in ‘loan/own’.
- This was in spite of the fact that word 1 was shorter when terminated by a voiceless stop, as opposed to being followed by a word-initial voiceless stop (plate ought/play taught).
 - Suggest that juncture cues are strong in these cases, duration cues may not be reliable in running speech.
- Judgments generally followed the transition portion going into word 2 - i.e. main juncture information was in onset of word 2.
- Final transition from word 1 was not informative, except in we loan/we’ll loan, two ran/tour an
 - onset/coda allophones of /l/, /r/.

Word boundary cues

Apparent cues:

- Glottal stop/laryngealization at the onset of vowel-initial words.
- Positional allophones of /l/, /r/, voiceless stops.
- These cues are segment specific, and therefore not available at every word juncture (e.g. you/th/read).
- Common to cite this paper and Lehiste (1972) as demonstrating that phonetic cues to segmentation are very limited.
- Worth investigating further. Intonational cues? (More on duration-related cues below).

Intonational cues to word boundaries in Korean

- Accentual Phrase
 - generally a lexical item plus a case marker or postpositions
 - marked by a melody: THLH (T=H if the AP initial segment is aspirated or tense, T=L otherwise)
- Kim (2003) found that subjects were better able to spot words in a prosodified stream of otherwise nonsensical syllables if the word begins an AP (rather than being medial in AP).

Lexical segmentation

- Segmentation via recognition runs into problems where words are contained within words.
 - E.g. ham, hamster; lip, tulip.
- Note that most work identifies these cases by examining broad segmental transcriptions. Not clear that this is always a good method for identifying embedded words (Pierrehumbert).
- Sequential models will find onset-embedded words (like *ham* in *hamster*).
- Trace is biased towards longer words because it is a spreading activation model:
 - the acoustic signal activate segment nodes, and these in turn activate the words that contain them.
 - The most active word is recognized.
 - Longer words contain more segments and thus receive more activation.

Embedded words

- Segmentation via recognition runs into problems where words are contained within words.
 - E.g. no, known; ham, hamster; lip, tulip.
- Note that most work identifies these cases by examining broad segmental transcriptions. Not clear that this is always a good method for identifying embedded words (Pierrehumbert).
- Sequential models will find onset-embedded words (like *ham* in *hamster*).
- Trace is biased towards longer words because it is a spreading activation model:
 - the acoustic signal activate segment nodes, and these in turn activate the words that contain them.
 - The most active word is recognized.
 - Longer words contain more segments and thus receive more activation.
- Do humans suffer from either of these problems?

Embedded words

- Gow and Gordon (1995) examined perception of two word sequences that are similar to a single word (*two lips/tulips*).
- Cross-modal semantic priming paradigm
 - Play word(s).
 - Present visual word for lexical decision.
 - Word is semantically related or unrelated to spoken stimulus
 - e.g. ‘tulips’ FLOWER, GRAMMAR.
- ‘two lips’ primes FLOWER, not significantly less than ‘tulips’.
- ‘two lips’ primes KISS as much as ‘warm lips’.
 - Suggests single word interpretations of two word sequences are activated in parallel with the two word interpretation.
- ‘tulips’ does not prime KISS.
 - Embedded words are not activated.
- Results imply that listeners do not rely on word boundary cues, but nor can they be relying on lexical strategies.

Embedded words

- How ambiguous are ambiguous phoneme strings?
- A variety of studies have shown that ‘embedded’ words are shorter than their freestanding counterparts, e.g. *ham-* in *hamster* is shorter than *ham*.
- This difference affects lexical activation (e.g. Salverda et al 2003).
- Crosswhite et al (2002) used eye-tracking to study perception of onset-embedded words.
- Subjects were instructed to manipulate pictures.
- Display included pairs of pictures corresponding to onset embedded words, like *ham*, *hamster*.
- Locations of eye fixations over time provide a good indication of the time course of lexical access.
- Found that subjects tended to look at the correct picture before the end of the first syllable.
- In a gating study, Davis et al (2002) found significant differences in the responses to embedded and containing words before the end of the first syllable, when total vowel duration is not known.

Embedded words

- Crosswhite (2002) found evidence that the formant trajectories of pairs like *ham* and *ham(ster)* differ. This could provide ‘early’ cues to the difference.

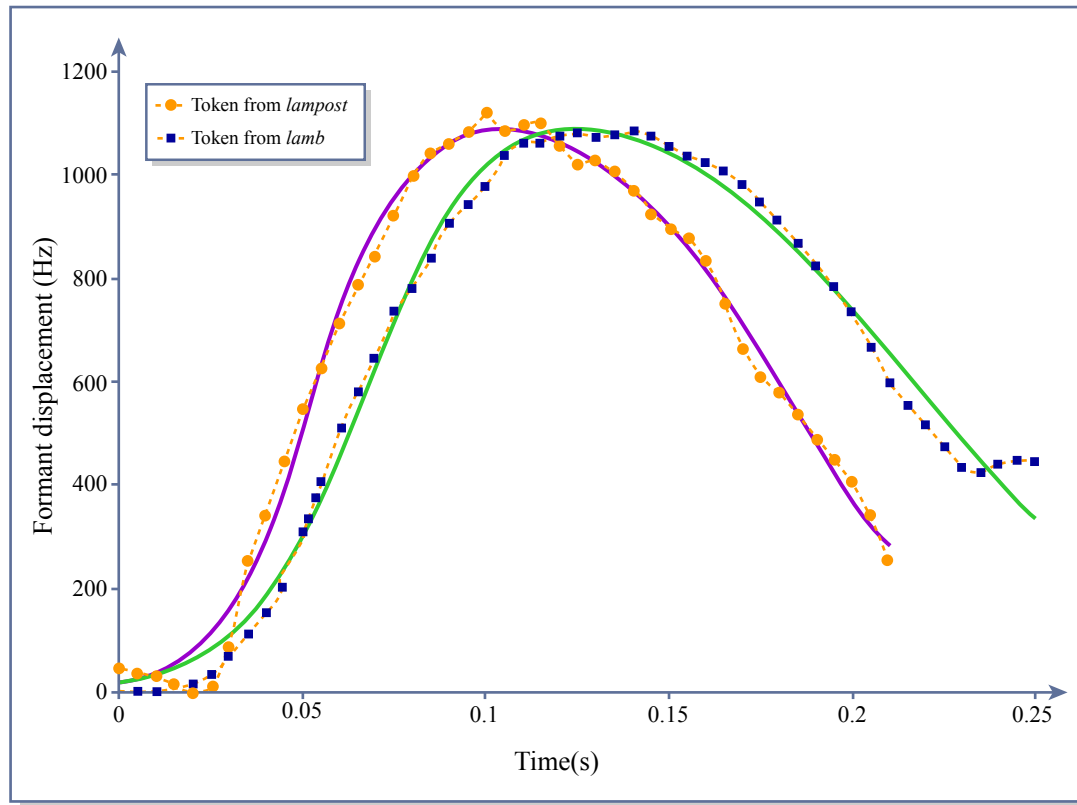


Image by MIT OpenCourseWare. Adapted from Crosswhite, Katherine. "Formant Trajectories in Onset-embedded Words." Manuscript, Center for Language Sciences, University of Rochester, 2002. (PDF)

Embedded words

- These studies indicate that there are ‘subphonemic’ differences between onset-embedded word pairs - i.e. there is less ambiguity than phonemic transcription suggests.
- The ability of listeners to use this information implies that it is not being discarded through construction of a segmental representation.
 - More detailed representation
 - Or differences in prosodic structure.

Embedded words

- Pierrehumbert (2001) has also emphasized that phonemic transcription is a bad guide to word-embeddings.

[Audio clip 1]

ham/ham(ster)

[Audio clip 2]

hamster

[Audio clip 3]

[Audio clip 4]

cat/cat(a)log

[Audio clip 5]

catalog

[Audio clip 6]

[Audio clip 7]

?

[Audio clip 8]

[Audio clip 9]

[Audio clip 10]

Voice recordings courtesy of Edward Flemming.

How often do we initiate lexical access?

- Models that posit segmentation via lexical access, with multiple candidate segmentations entertained in parallel (e.g. TRACE) must address the question of how often we initiate lexical access.
- TRACE assumed an intermediate, segmental representation, with lexical access initiated at each phone.
- But what if the intermediate representation is more detailed. There is a variety of evidence that sub-phonemic detail influences lexical access, including the results just reviewed.
- So do we initiate lexical access every 10 ms? This could lead to a huge number of word hypotheses, and competition between slightly staggered versions of the same word (Pierrehumbert 2001).
- Pierrehumbert's solution: pre-lexical phonological processing, not to abstract away from phonetic detail prior, but in order to identify possible word boundaries. Phonetic detail may be retained for lexical access.