What is special about prosody in processing (across languages, varieties or nativeness)?

Anne Cutler

ARC CENTRE OF EXCELLENCE FOR
THE DYNAMICS OF LANGUAGE

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS

the MARCS institute

WESTERN SYDNEY UNIVERSITY

What are we – prosodists?

- Prosody is the linguistic dimension of which language users are least aware
- “Around the edge of language” (Bolinger)
- Few languages choose to incorporate it (in any way) in their writing system
- Prosody of L1 isn’t taught (maybe for the above reason?)
- And this carries through to L2 teaching
- Cross-language differences are (thus?) not obvious
- Also overlooked (perhaps again for the same reason?): influence of vocabulary structure on prosodic processing

Relatively underexplored topics in prosody

1. How to teach the use of prosody

2. Using prosodic data to explain effects in other domains

3. Prosodic effects of ingesting alcohol

4. Cross-language comparisons of prosodic processing

   (Why has there been relatively little of this?)

Teaching sentence prosody in L2 listening

Today we've discussed the development of Facebook, tomorrow we'll discuss:

(1) *the impact of Facebook*

(2) *the development of Twitter*

(3) [can't tell!]

(Reed, 2014)

Native listeners automatically use prosodic information (such as pitch accents to mark emphasis in English).
But this is never taught! So L2 listeners often overlook it.

Prosodic data explaining effects in other domains

e.g.: Cross-talk between levels of (speech) processing?

1980s: Is it more likely for speech errors to create words (*map task -> tap task*) than nonwords (*mid term -> tid term*)?

Some errors are clearly word-level (*left -> right, green -> yellow*) vs. sound-level (*long-short -> shong-lort*).

We know: word-level errors are more likely to be corrected and the correction will be more emphatic.

So are errors like *map task -> tap task* word or sound errors?

Prosodically, they are corrected like sound errors… so most probably they are just real words by accident.

Thus speakers' prosody in correcting errors can illuminate speech errors' source. (Shattuck-Hufnagel & Cutler, *ICPhS* 1999)

Cross-language comparison of prosodic processing

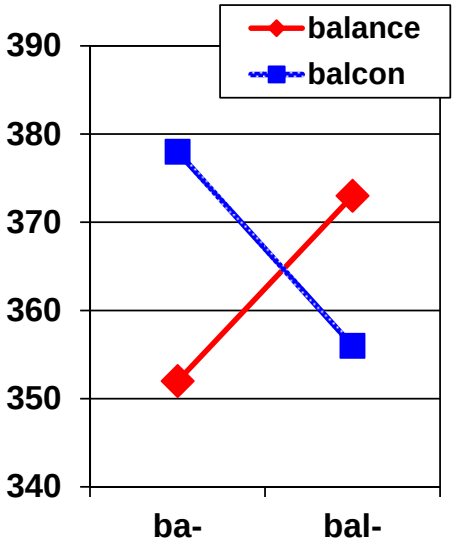
Growth areas:

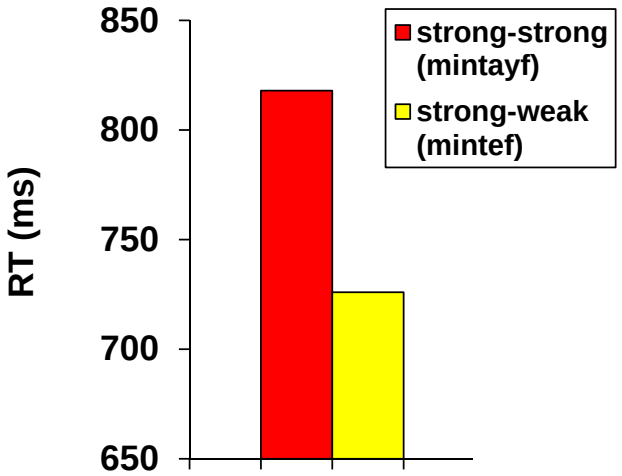
- Prosodic structure in languages where it was not previously described
- Prosodic processing of many kinds
  - Word level
  - Sentence level
- in many languages (even unto those where “word” and “sentence” level are not easily distinguished)

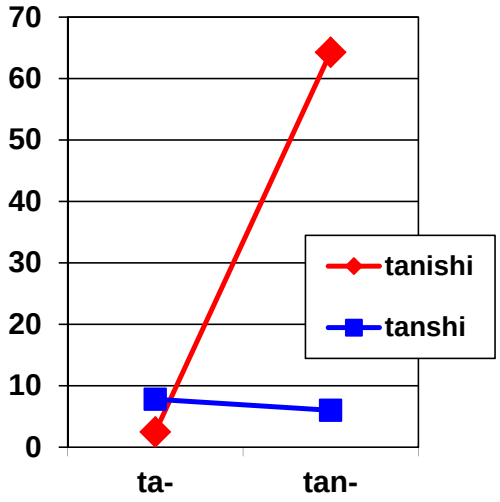
So why not predict where structures affect processing?

Speech segmentation: Language-specific rhythm induces language-specific processing

	1. French:	2. English:	3. Japanese:
RHYTHM:	syllabic	stress	moraic
SEGMENTATION:	syllable-based	stress-based	mora-based





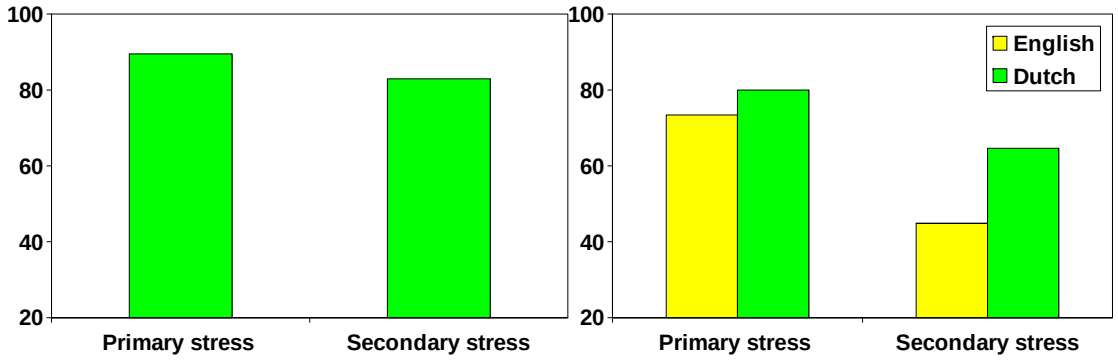


Target=syllable =faster	Word=cross stress boundary = slower	Target=mora =faster

(Mehler, Dommergues, Frauenfelder & Segui, 1981; Cutler, Mehler, Norris & Segui, 1983, 1986, 1989, 1992; Cutler & Norris, 1988; Otake, Hatano, Cutler & Mehler, 1993; Cutler & Otake, 1994)

Word level: Adult use of stress cues in English, Dutch

- English and Dutch: similar in stress phonology
- If 2 words begin with same segments but differ in stress:
  - CARton / carTOON; CASHew / cashIER; MUsic / muSEum
- How quickly can we tell what word we're hearing?



voor- from *VOORnaam* *'first name'*

voor- from *voorNAAM* *'respectable'*

mus- from *music*

mus- from *museum*

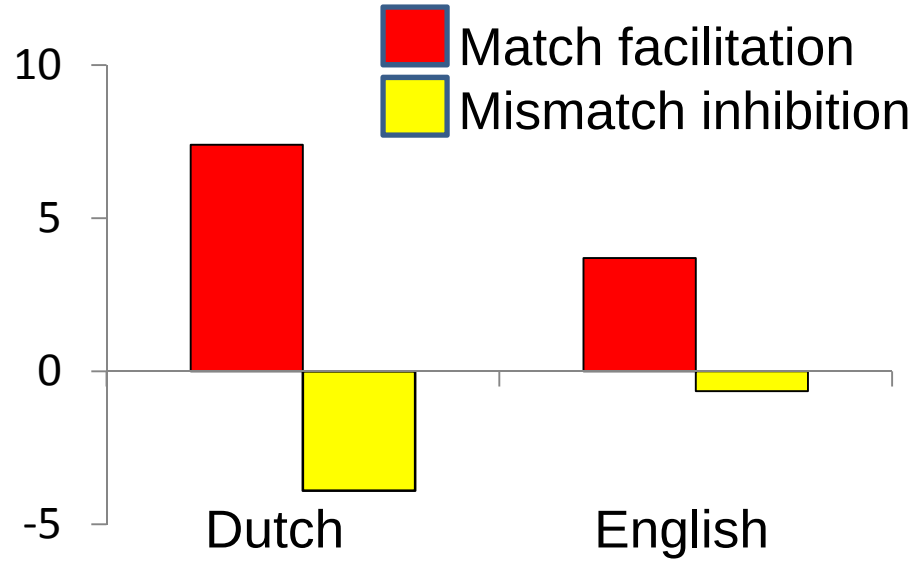
(Cutler & Donselaar, 2001; Cooper, et al., 2002)

Dutch listeners do this well in Dutch – and also in English! In fact, so well that they outdo English listeners (who in the case of secondary stress do no better than chance).

Stress cues in word recognition

- Primes: neutral sentences ending in initial bisyllabic portions of word pairs contrasting in stress:
  - *domi-* from *DOminee* or *domiNANT* (Dutch)
  - *admi-* from *ADmiral* or *admiRAtion* (English)
- Recognition of visual word (e.g. ADMIRAL) compared after prime match (*ADmi-*), mismatch (*admi-*), or control
- Match facilitation should always occur (segmental match alone ensures that)
- Mismatch inhibition is the crucial evidence that stress cues are actually being used

(Donselaar et al. 2005; Cooper et al., 2002)

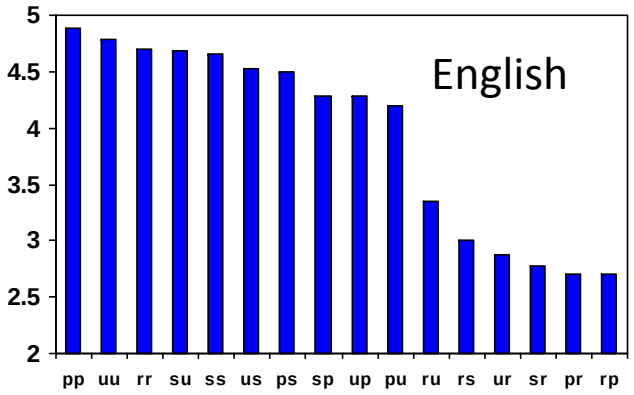


Sensitivity to levels of stress

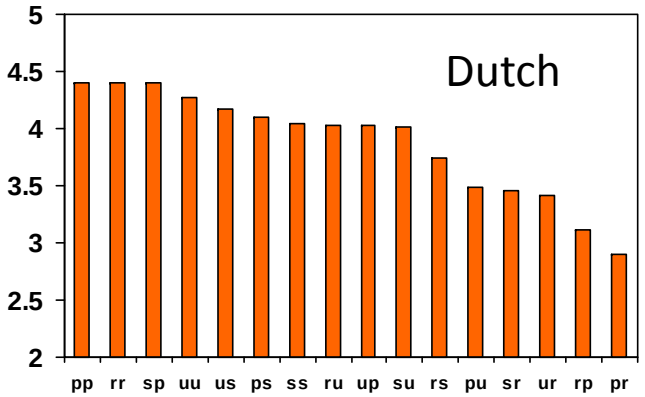
Cross-splicing study
(Fear, Cutler & Butterfield, 1995)

autumn) Within each set of 4 words, all
automation) 4 initial vowels attached to all
automata) 4 word bodies. Listeners rated
atomic) the spliced forms for naturalness.

Ratings for spliced words



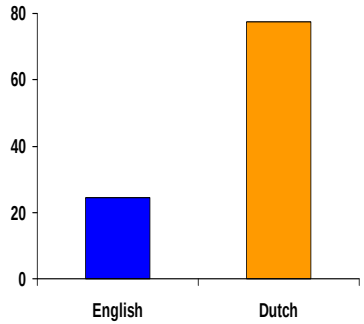
English listeners only care about vowel quality. Ratings for cross-splicings between *autumn, automation, automata* are not significantly different from identity-spliced words.



Dutch listeners produce a more graded response to the same English stimuli

(Cutler, 2009)

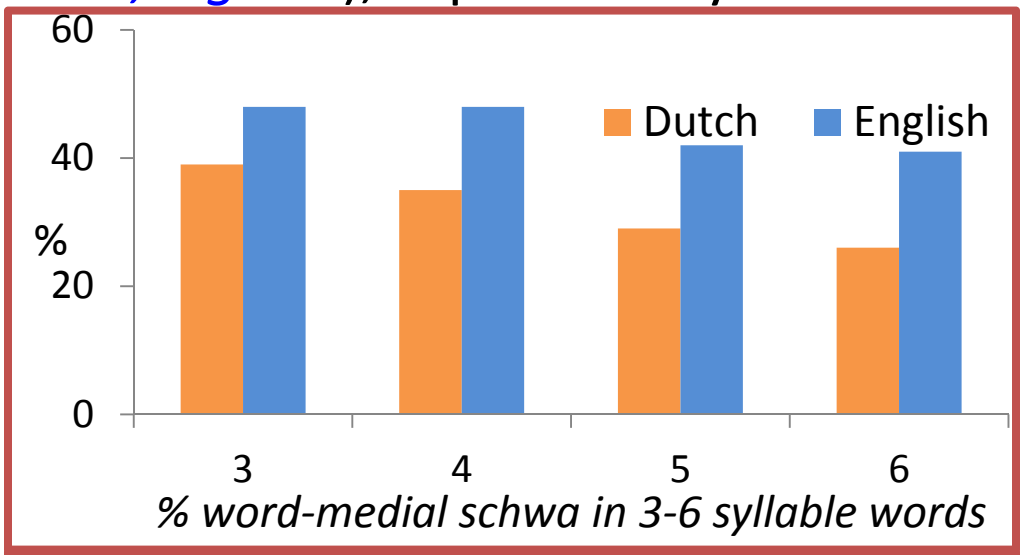
Unstressed vowels replaced by vowels with secondary stress: percent decrease in rating

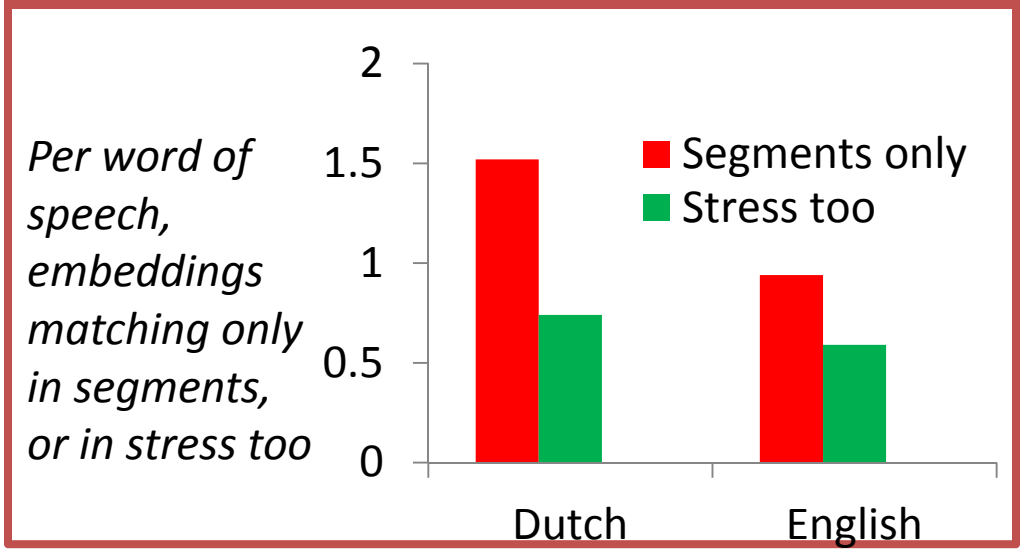


The reason for a cross-language difference

- English vs. Dutch: stress rules are similar, but English more often reduces unstressed syllables (*cobra, cigar*...), esp. medially (*octopus, dominant*).
- To choose among word candidates in English, segments usually suffice
- Dictionary comparison of number of embedded words
  - segmental match only
  - segments + primary stress:
- Dutch vocabulary gives greater payoff for use of suprasegmental cues

(Bruggeman & Cutler, 2016; Cutler & Pasveer, 2006)





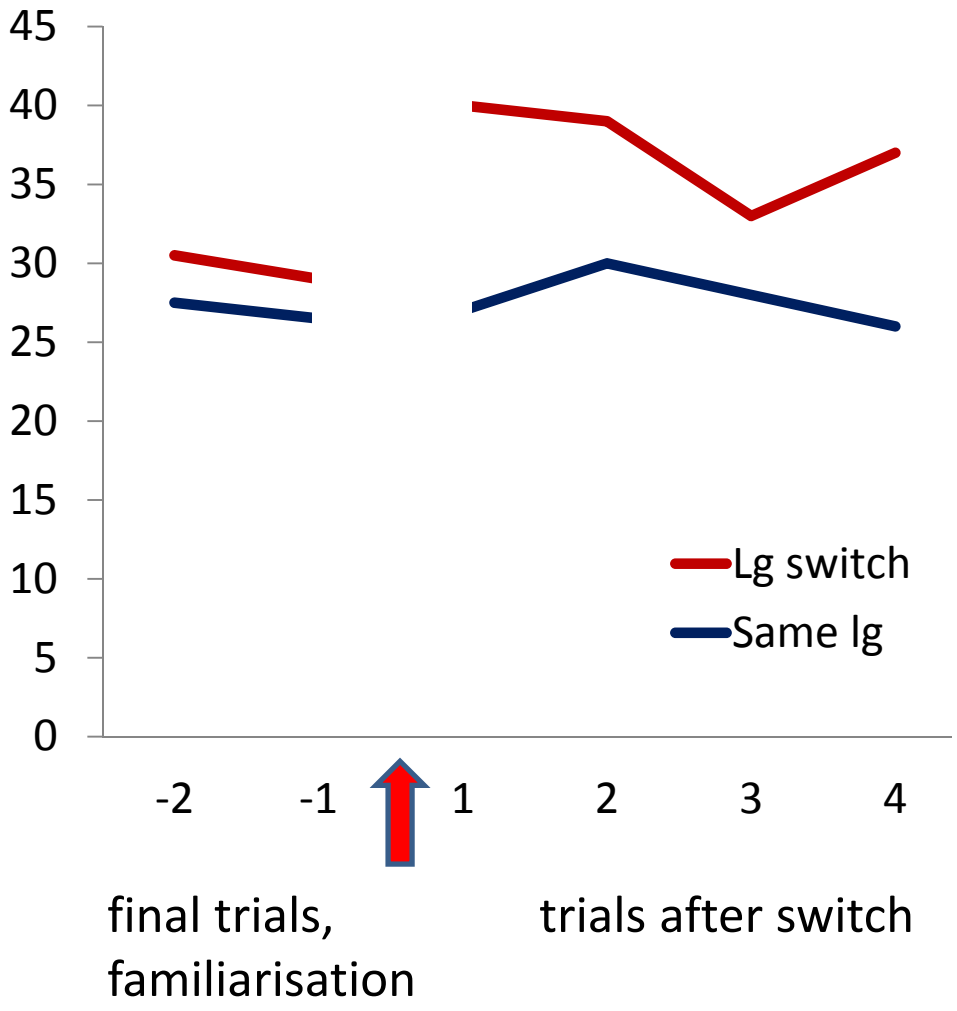
This prosodic sensitivity starts really early

Rhythmic structure of the native language is part of infants' auditory experience before birth

Newborns prefer native rhythm

and (at two to four days) detect a switch across languages that differ in rhythm but not across languages with the same rhythm (Nazzi, Bertoncini & Mehler, 1998)

They still fail to discriminate the native language from another with the same rhythm at 2 months (Christophe & Morton, 1998)



Early sensitivity to acoustic correlates of stress

BAba vs. *baBA*?

Yes, at 2 months for English-learners (Jusczyk & Thompson, 1978).

Yes, at 6 months and at 9 months for French- and Spanish-learners (Skoruppa et al., 2009; 2013)

LApi, NAku, Tlla vs. *laPI, naKU, tiLA*?

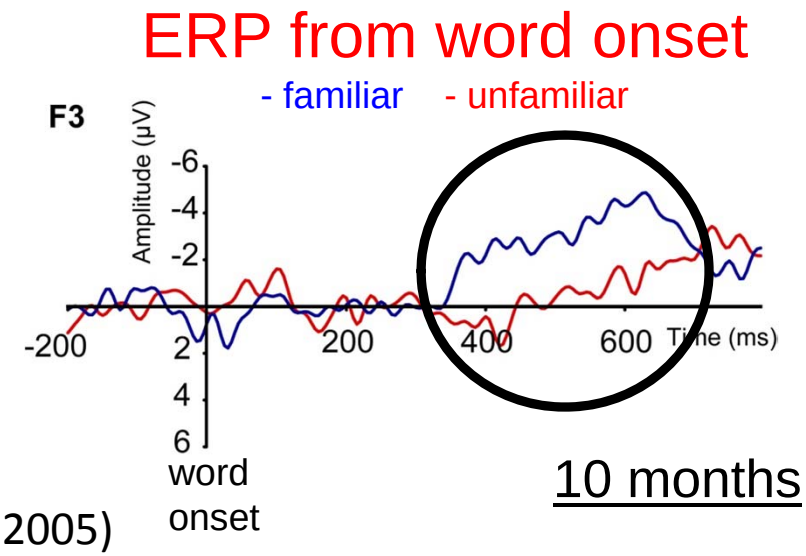
Yes, at 8-12 months for English- and Spanish-learners, but not for French-learners (Skoruppa et al., 2009; 2011)

Vocabulary characteristics?

By 9 months they know the typical native-language word shape (in English: trochees) (Jusczyk, Cutler & Redanz, 1993)

Infant speech segmentation (ERP studies)

- Behavioural studies (headturn task) show that infants can recognise a familiarised word if it recurs in a sentence (i.e., segment speech).
- A clear speech segmentation effect also is seen in ERPs (Event-Related Potentials: the infant brain's response to speech stimuli).
- Familiarised words recurring in sentences elicit a more negative response than unfamiliar words.
- The target words were all trochaic (*mosterd, hofnar*...)



10 months

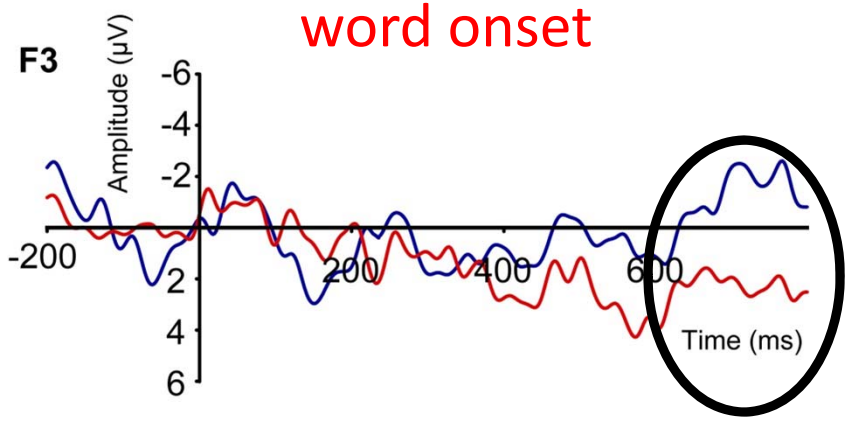
(Kooijman, Hagoort & Cutler, 2005)

Metrical segmentation at 10 months

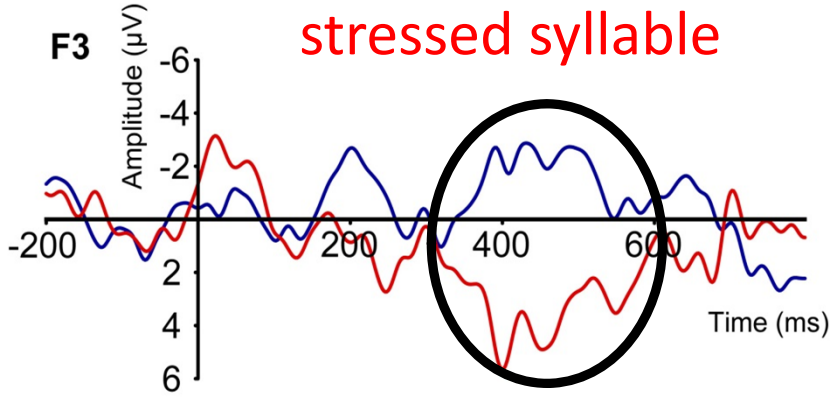
- Behavioral studies in English showed that segmentation is driven by stressed syllables.
- This is also true of Dutch 10-month-olds' brain responses (ERPs).
- In weak-strong words (*getij*, *legaal*), just as in strong-weak words (*tijger, galig*), the typical negativity is triggered by the stressed syllable.
- Infants' segmentation of Dutch also driven by stress

(Kooijman, Hagoort & Cutler, 2009)

WS: ERP keyed to word onset



WS: ERP keyed to stressed syllable



- familiar - unfamiliar

Infant use of stress cues: English, Dutch

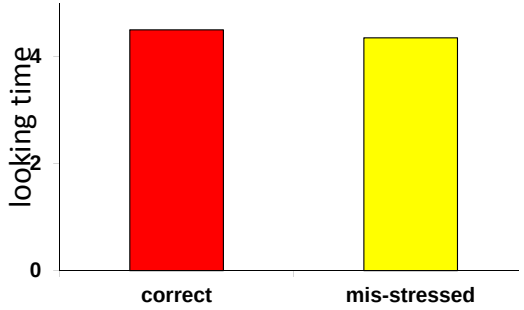
- The picture from infancy: stress differences are easily perceptible for infants exposed to stress languages
- Distributional asymmetry of stress in the lexicon is picked up early (in fact, in infant input, asymmetries tend to be exaggerated):

Lexical Words:	S-initial	W-initial
English lexical data	.902	.098
English infant input	.968	.032
Dutch lexical data	.883	.117
Dutch infant input	.971	.029

- As a result, earliest word segmentation skills exploit the stress patterning much as adult segmentation does
- English and Dutch development appears very similar

Mis-stressing effects

- One symptom of English-speakers' segment-only lexical processing is that mis-stressing is overlooked if segments are unaltered (Bond, 1981; Cutler & Clifton, 1984; Cutler, 1986; Slowiaczek, 1990).
- When does this appear? NB familiarity effects for words at 11 months but not at 9 months (Hallé & Boysson-Bardies, 1996).
- For 11-month-old English-learners, familiar words are as good mis-stressed (*thankYOU, baBY, TOnight*) as correctly stressed (*THANKyou, BAby, toNIGHT*; Vihman et al., 2004).
- Segment changes (e.g., *VAby*) are treated like unknown words (Swingley & Aslin, 2002).

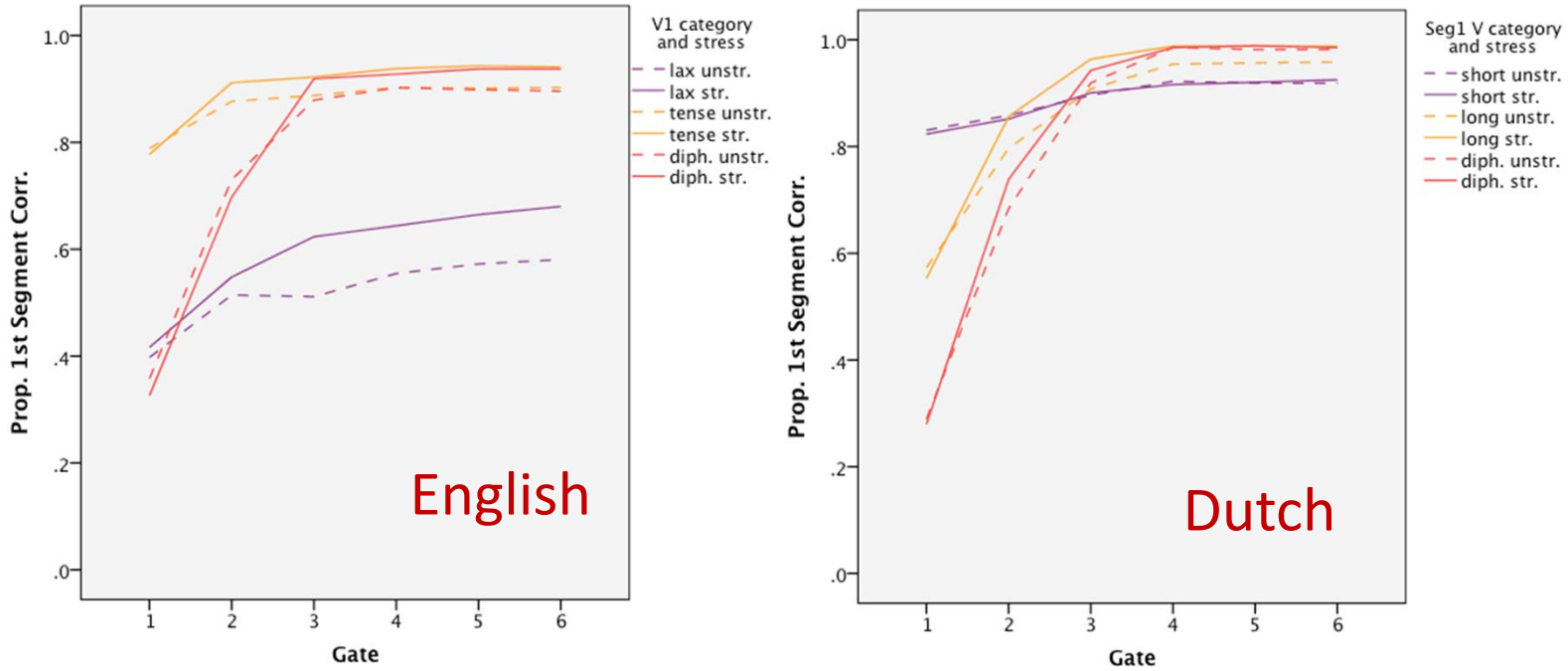


- So as soon as an initial lexicon is in place, English-learners give more weight to segments than to stress.

Consequence: English/Dutch vowel identification

Evidence from 2 giant speech perception experiments

In each case: All possible diphones of the language, heard in fragments



- Solid lines: stressed vowels. Dashed lines: unstressed vowels.

English: big effect (They don't expect vowels in multiple stress versions)
Dutch: little stress effect (They are used to differently stressed vowels)

Smits, Warner, McQueen & Cutler, 2003; Warner, McQueen & Cutler, 2014; Warner & Cutler, 2017

Using prosody in listening

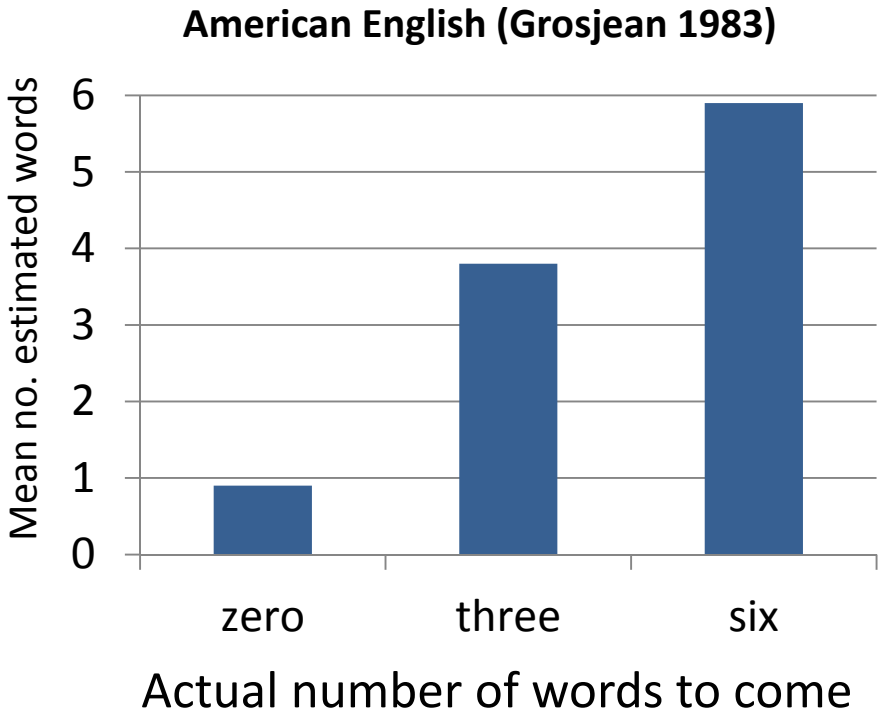
- Use of prosody in listening starts early
- And at the word level is driven by the lexical structure – apparently from the earliest opportunity…
- At higher levels of processing this should presumably also be the case
- (Missing data here? – language-specificity in children's earliest processing of prosody)
- Even language-specificity in adult prosodic processing is understudied!

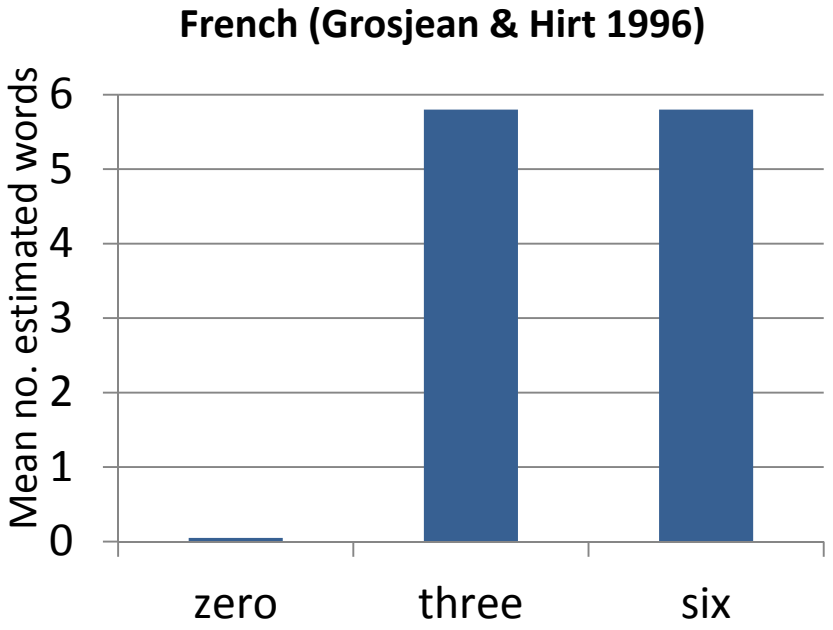
Sentence level: Using prosody in listening

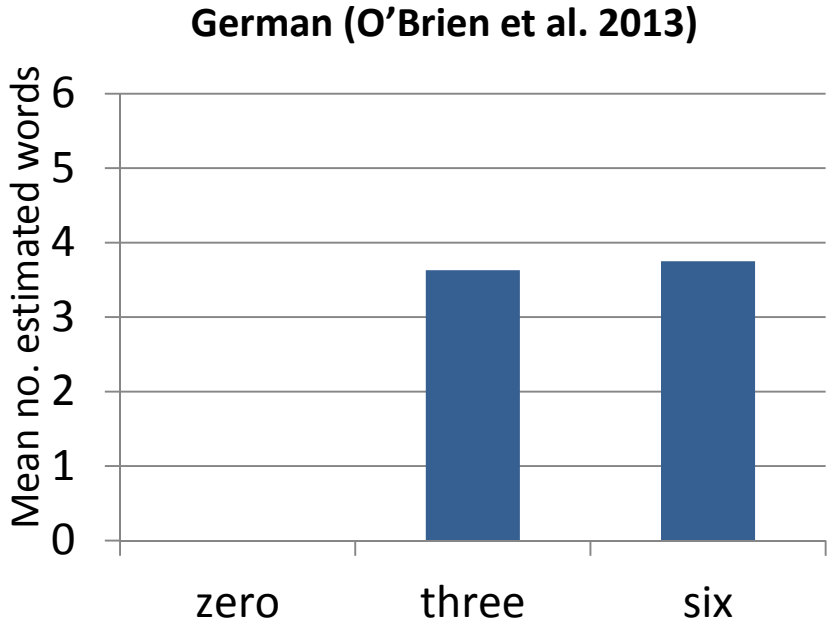
- Native English listeners use prosody in many ways. E.g. predicting how long an utterance is going to be:
- *Yesterday James fixed the lock on the door of his car*
- Prior gating experiments have shown:
- At the first phrase boundary (*lock*) American English listeners can tell whether 0 or 3 or 6 words are coming
- French listeners (with French) can tell 0 from more (3/6)
- German listeners (with German) can tell 0 from more
- L2 listeners (English L1) to German can tell 0 from more

(Grosjean, 1983; Grosjean & Hirt, 1996; O'Brien, Jackson & Hendricks, 2013)

English listeners' sensitivity to upcoming utterance length outpaces sensitivity in French and German







Using sentence prosody in listening

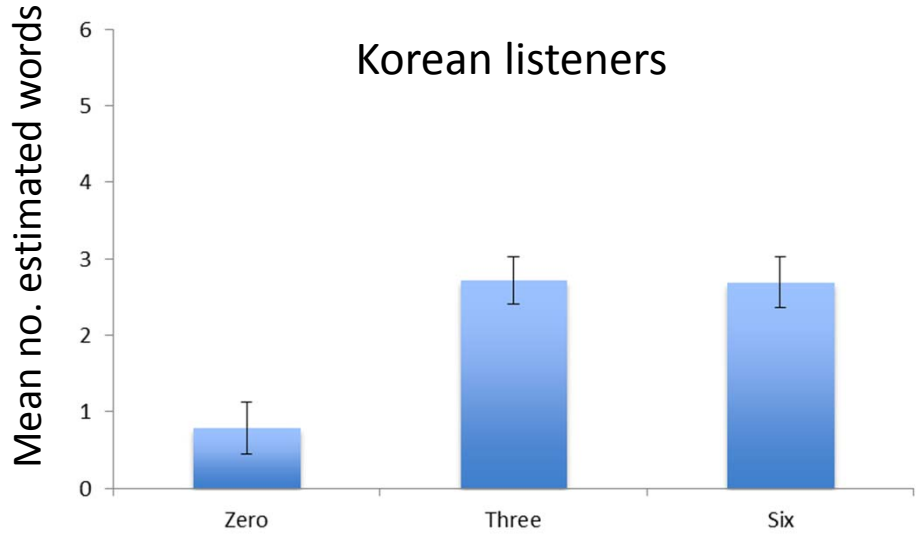
- English may offer more cues to utterance length
- Do L2 listeners differ in how well they can use the cues?
- E.g. Korean – an *edge-prominence* language
  - L1 requires attention to phrase boundaries, so listeners may be able to use these subtle phrase boundary cues in their L2 English?
- Or Vietnamese – a *tone* language
  - L1 requires pitch contour processing for word recognition, so using pitch in a different way (for predicting utterance length) may be really hard?

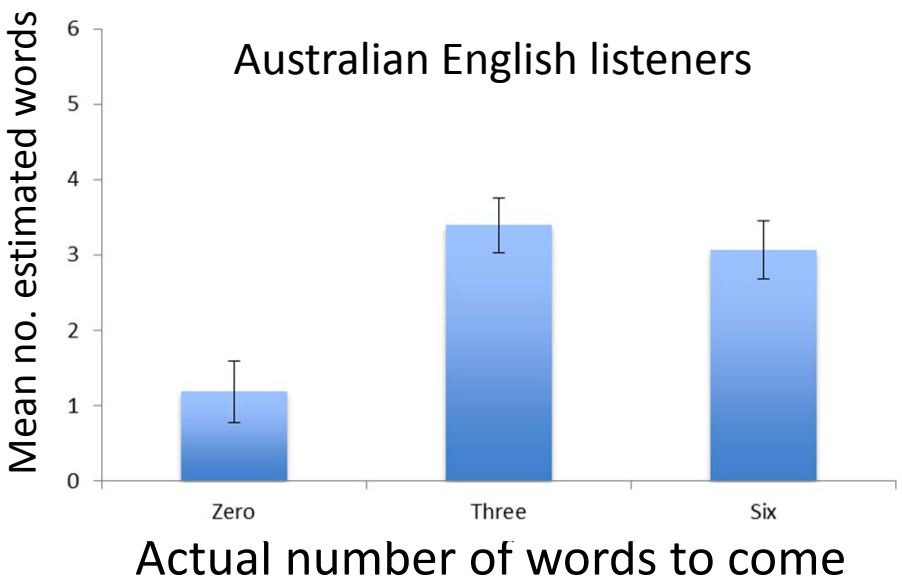
Using sentence prosody in listening

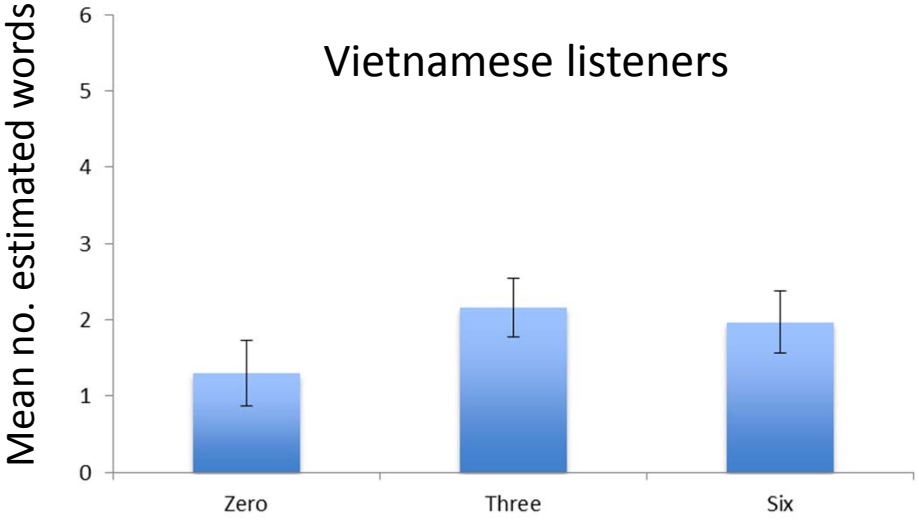
- Our study: 20 each Korean, Vietnamese & AusE listeners
- L2 listeners had high English proficiency (*LexTALE*)
- All materials in Aus English
- 33 sentences, recorded in short, medium, long versions
  - *Yesterday James fixed the lock on the door of his car*
- Only the initial portion (e.g. up to *lock*) presented
- …in 8 fragments (incremental gates) per sentence
- Listeners decided for each fragment whether the sentence was the short, medium or long version, and rated their confidence in that choice

(Jeske, Kember & Cutler, 2017)

Results: Australian English sentences







Language: n.s.
Interaction: n.s.
Sentence length: $p < .001$
0 vs. 3, 0 vs. 6: $p < .001$
3 vs. 6: n.s.

How long is an Aussie sentence?

- Answer: It's either finished or it's not
- Korean and Vietnamese with L2 English produced the same results pattern (also seen with French & German)
- And so did the native AusE listeners!
- AmE listeners in the 1980s discerned a difference that 21st century AusE listeners can't. Why?
  - **The materials** (AusE never had the same cues as AmE? The language has changed? Or: AusE has "uptalk"?)
  - **The participants** (1980s kids listened better than kids do now, or Boston kids listen better than Sydney kids?)
  - **Something else** (e.g. the 1980s results were flawed?)
- What's the next step (if any)?

(Jeske, Kember & Cutler, 2017)

What is special about prosody in processing?

- Prosody is special – in the extent to which it is ignored
- (Cross-language processing studies of prosody maybe just need more good examples to follow?)
- What can we do about making prosody more obvious?
- Can we go beyond our current options:
  - Drawing contours etc.
  - Measuring suprasegmental dimensions of speech components
  - Labelling specified prosodic events (ToBI)
- i.e., can we get a better way of presenting prosody? – one which is transparent and accessible to non-initiates, is not function-driven but neutral