

Experiment on Acoustic and Phonetic Details

Rolf Carlson

This presentation is based on ...

Rolf Carlson and Sarah Hawkins (2007) *When is Fine Phonetic Detail a Detail?* ICPHS 2007

Rolf Carlson (2007) *Using acoustic cues in stop perception.* Fonetik 2007.

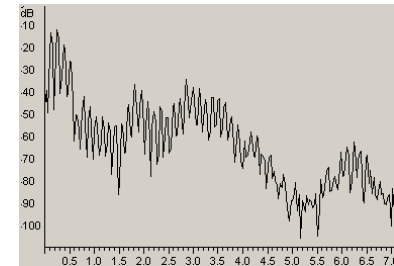
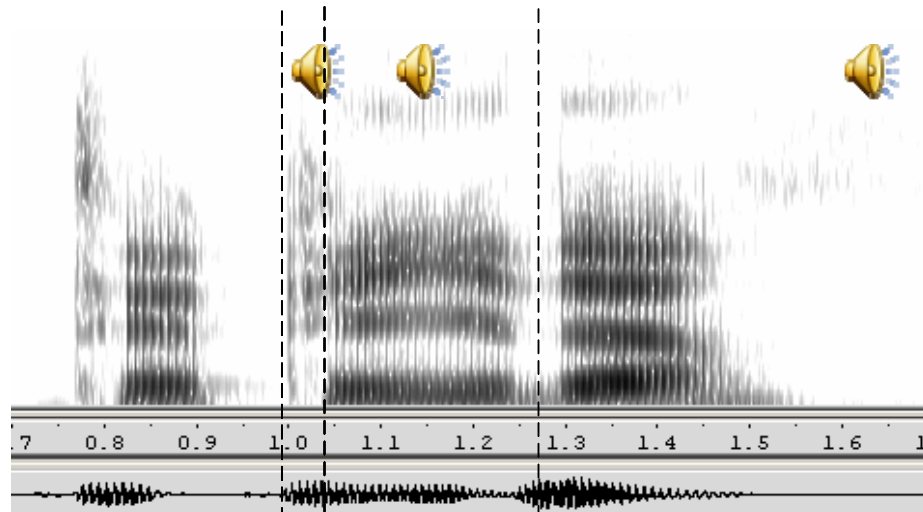
Background for discussion

- Experiments exploring the relation between acoustic cues and perceived phonetic units
- ... to illuminate and support the hypothesis that speech perception is a **dynamic and adaptive** perceptual process in the use of fine phonetic details.

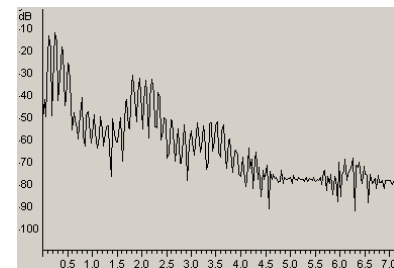
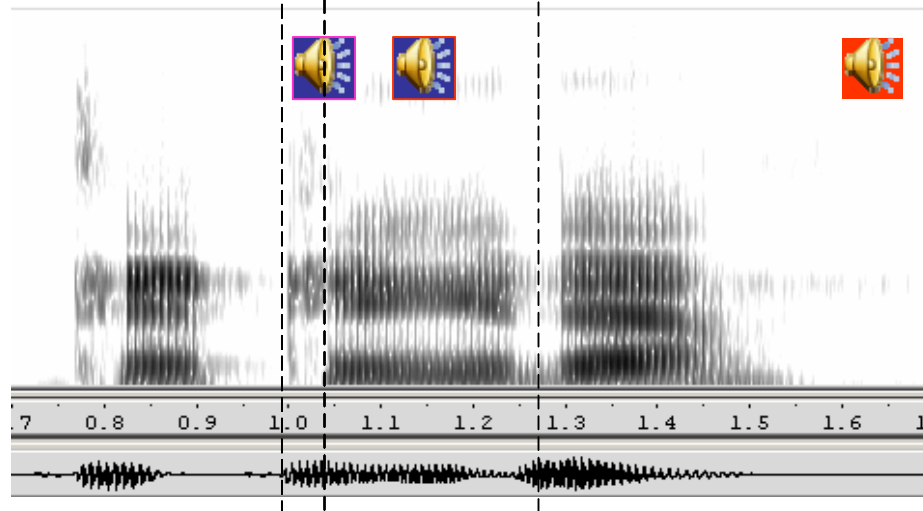
Active processing in speech perception

- Unpublished experiment, carried out during the 70s by Gunnar Fant.
- Third formant in **one** natural front vowel was moved with the help of a pole-zero filter resulting in a **perceptual vowel shift**.
- When the same transformation was applied to a **sequence** of different vowels, perceived vowel identities **did not shift**.
- The perceptual process was able to identify the filtering as a distortion and disregard that a formant was misplaced.

Perception Experiment Again



filter: pole 2200 Hz (150 Hz bw) zero 2850 Hz (200 Hz bw)

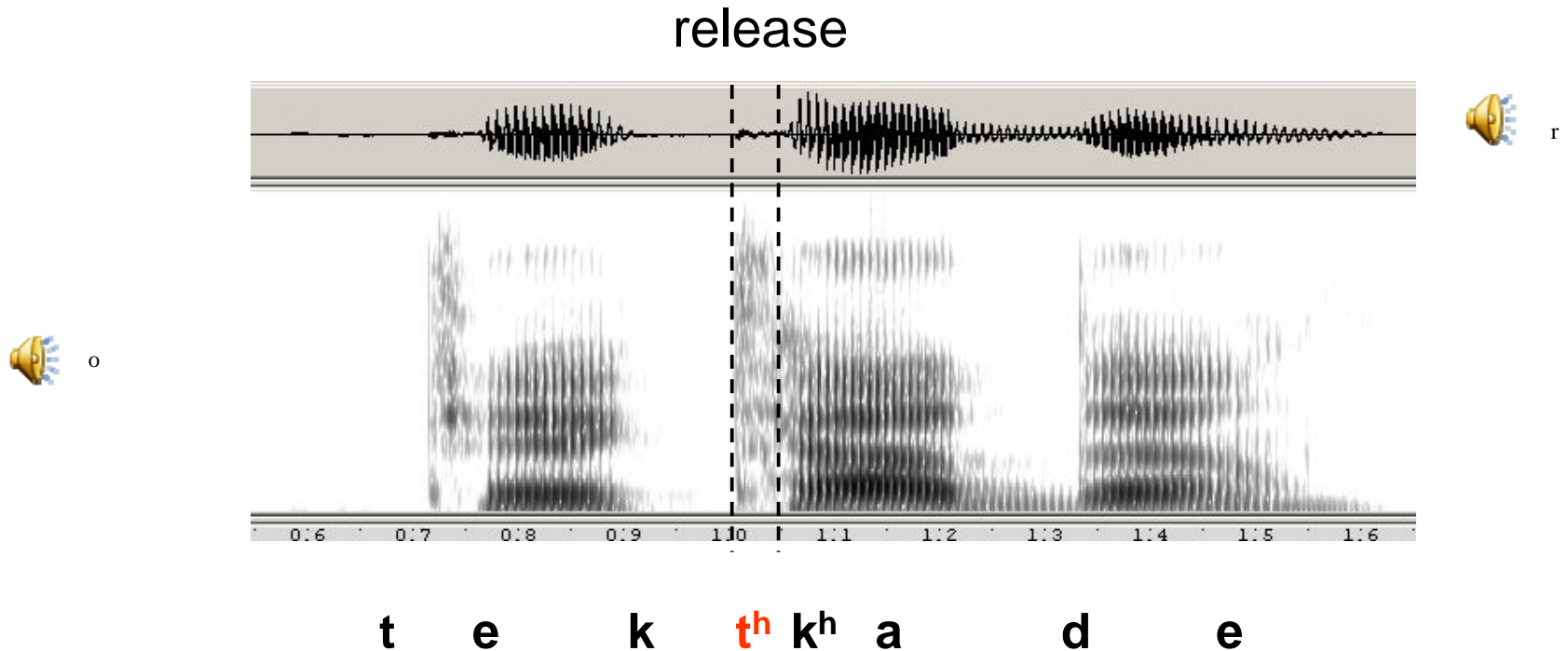


Stop Experiment

- A straightforward waveform splicing technique has been used to create stimuli with sometimes contradicting acoustical cues.
- 18 nonsense words /te'CVde/ (Swedish speaker)
 - C one of three voiceless stops /p t k/
 - V one of six vowels /a a: i i: u u:/
- 7 subjects
- five types of stimuli
- *release* stimuli
 - 40 ms from the stressed C release was replaced by the equivalent segment from another C stop.

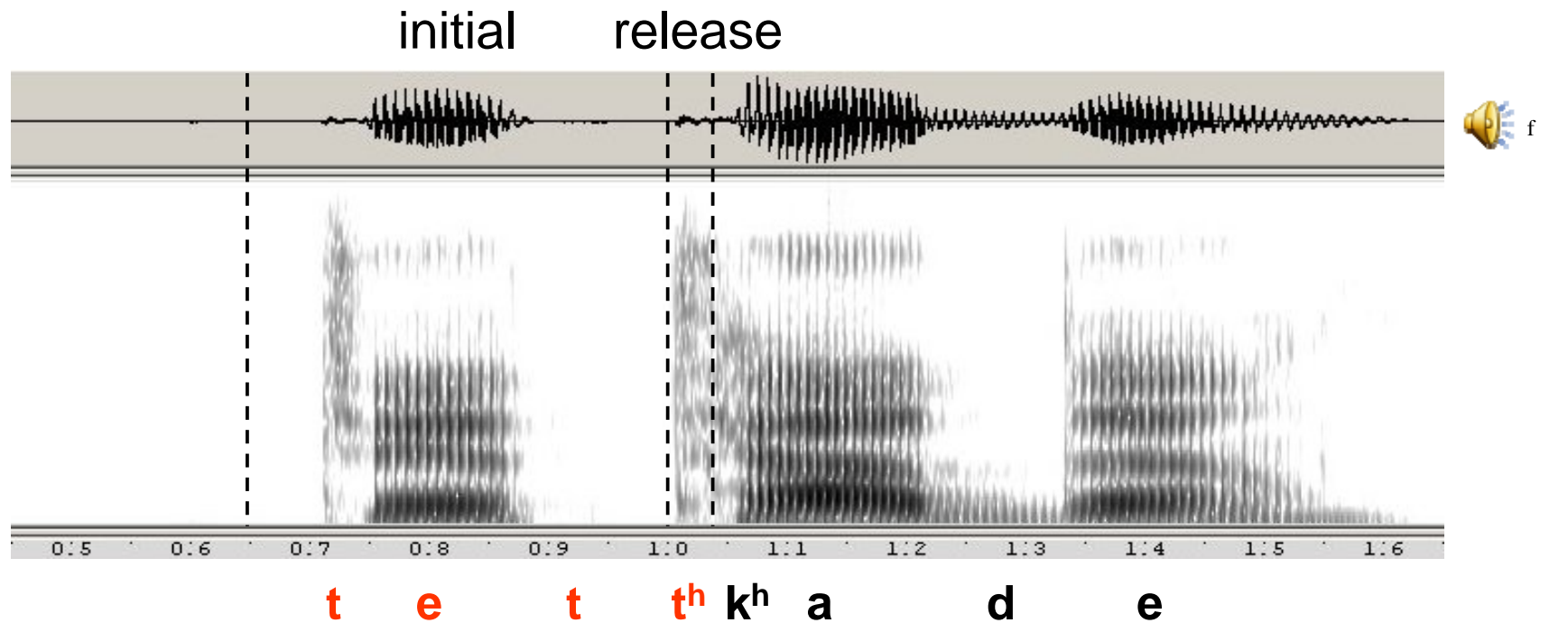
release stimuli

- Example of a *release* stimulus.
The **stressed stop release** (40ms), from "te'**t**adde"
has replaced the corresponding part in "te'**k**adde"



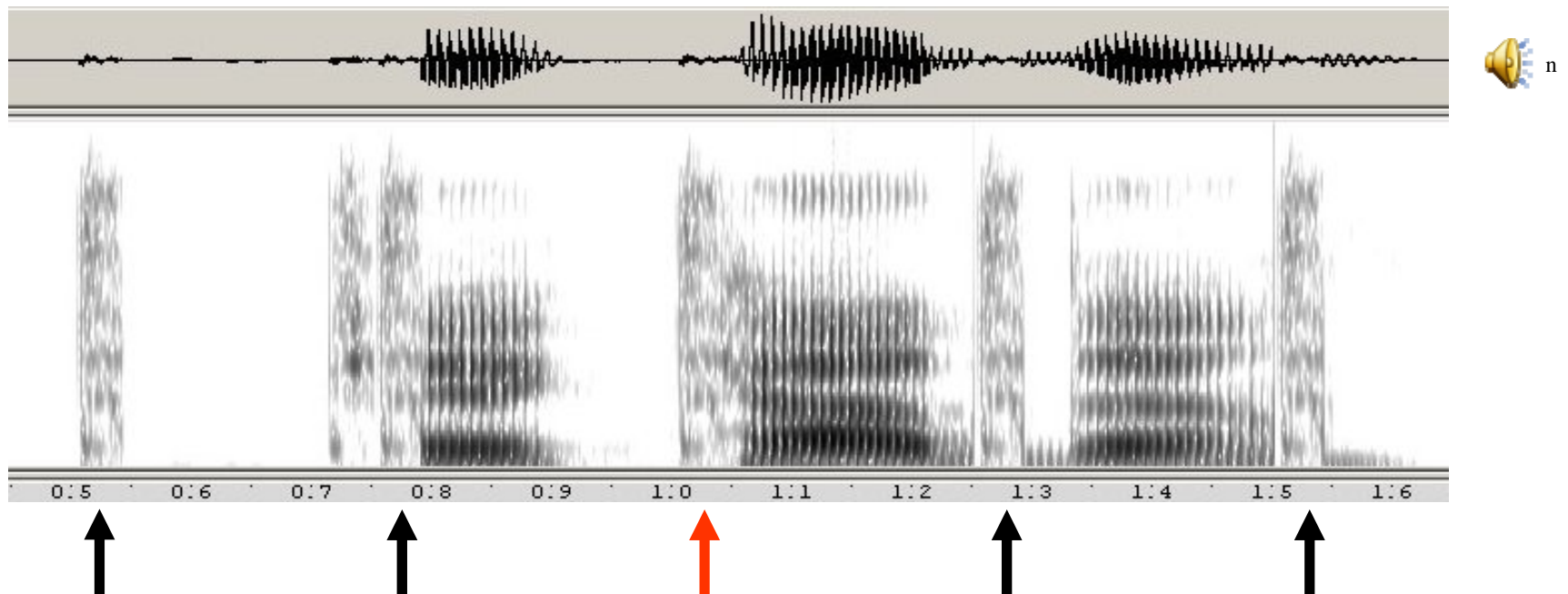
initial+release

- Example of a type *initial+release* stimulus.
The **initial syllable and the stop release** (40ms), from "te'tadde" has replaced the corresponding part in "te'kadde"



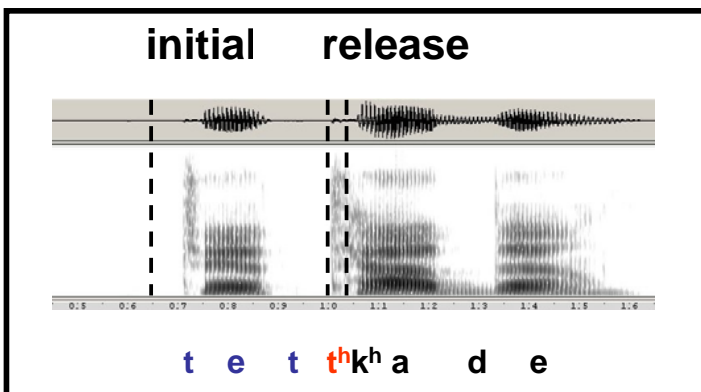
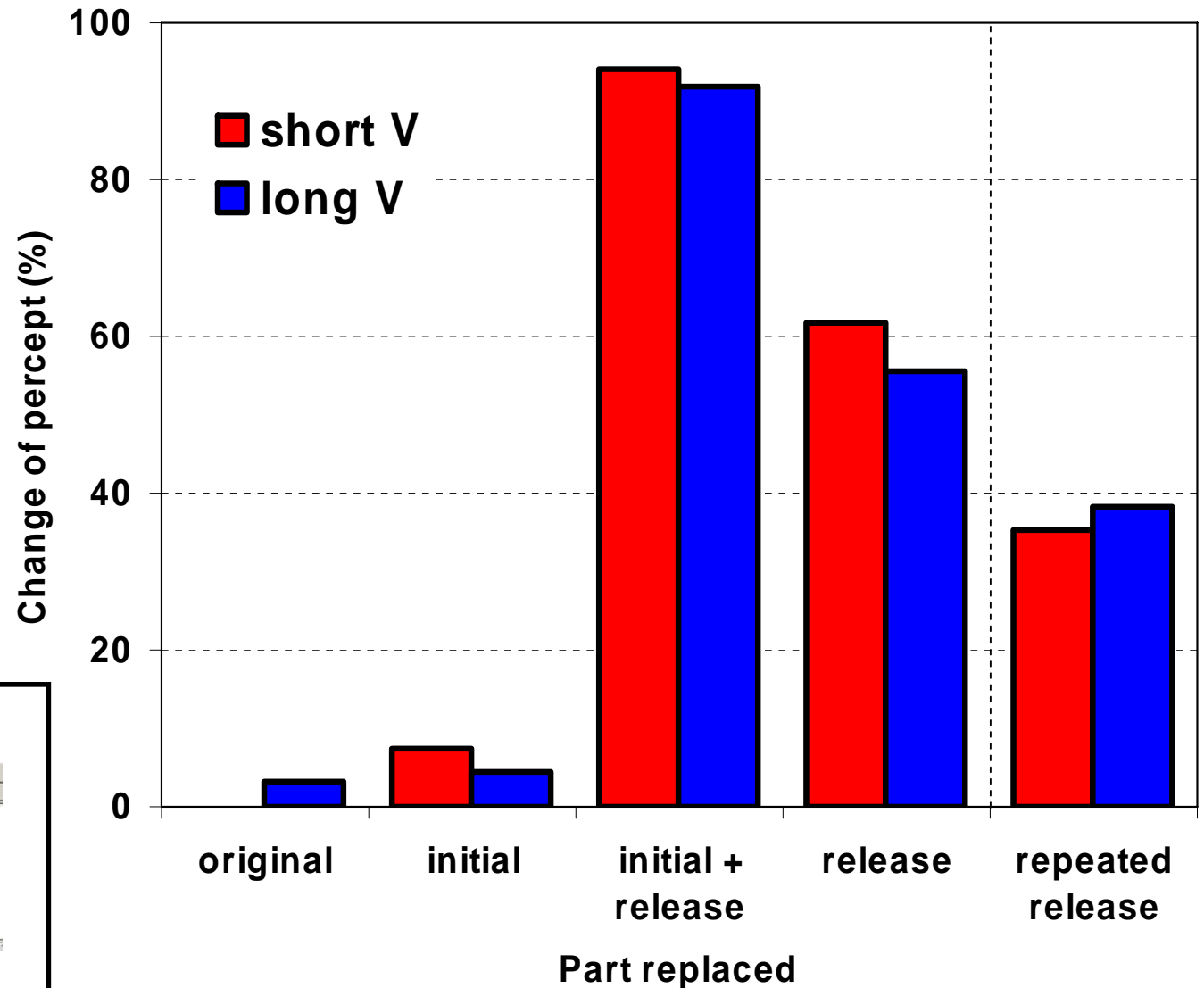
repeated release stimuli

- Example: The stressed release (40ms) in "te'tade" has replaced the release in "te'kade" (red arrow). And, this same t-release is repeated at regular intervals (marked with black arrows).



Results

Change of percept (COP) grouped according to stimulus type and vowel length



Possible Lessons

- A temporal distortion can have the same kind of impact on a percept as a spectral distortion.
- The fine phonetic details have reduced importance for the classification if they are interpreted as potential disturbance.
- The cues in the preceding vowel are weaker than cues in the release. But the combination of cues in the preceding vowel and the release generates a stronger COP than a simple addition of the separate cues (93% > (6+59)%).
- When even weak cues are coherent, perceptual decisions are more consistent.
- This underlines the importance of long-domain integration of acoustic information, and the central role of perceived context in this process.

World knowledge

- Listeners' world knowledge dictates the perceptual salience of an attribute of the signal in different ways depending on circumstances.
- The processing requires integration of information over some considerable time.
- Perceptual decisions must be context-sensitive representations of certainty.
- Determining what is context, and what is signal, is presumably partly inherent to the perceptual system, and partly a function of individual experience, current expectation, and attention.
- Perceptual decisions about speech will be very sensitive to signal properties and task demands
- This long-domain property of perceptual decisions poses a real challenge for computational modelling of speech recognition by humans and machines.

When is an Fine Phonetic Detail a Detail?

- A temporal distortion can have the same kind of impact on a percept as a spectral distortion.
- The fine phonetic details have reduced importance for the classification if they are interpreted as potential disturbance.
- When even weak cues are coherent, perceptual decisions are more consistent.
- This underlines the importance of long-domain integration of acoustic information, and the central role of perceived context in this process.
- Determining what is context, and what is signal, is presumably partly inherent to the perceptual system, and partly a function of individual experience, current expectation, and attention.
- This long-domain property of perceptual decisions poses a real challenge for computational modelling of speech recognition by humans and machines.
- We have tried to make clear that FPD is not all about tiny details, and is not all fine.