AMSTERDAM CENTER FOR LANGUAGE AND COMMUNICATION

ACLC

Timing of Turntaking: Early Responses and Use of Intonation in an Elicited Minimal Response Task

Wieneke Wesseling Rob van Son

ACLC
Phonetic Sciences
University of Amsterdam

Dag van de Fonetiek 2005

Introduction: Motivation

In understanding language, different sources of information are used:

- syntactic information
- semantic information
- visual cues (e.g. gaze direction, gestures)
- prosodic information (loudness, duration, tempo, pauses, pitch)

Main Question:

What is their relative importance?

Introduction: Motivation

In understanding language, different sources of information are used:

- syntactic information
- semantic information
- visual cues (e.g. gaze direction, gestures)
- prosodic information (loudness, duration, tempo, pauses, pitch)

Main Question:

What is their relative importance?

Introduction: Task

Minimal Response Task:

Identification of TRP's in Dialogue

- Reaction Time (RT) task
- Identify when to start speaking
- by saying 'AH'
- more 'natural' task than pushing button
- responses recorded with laryngograph

Assumption: at this point there is recognition of (at least part of) the previous utterance

Introduction: Questions

Questions adressed in this talk:

- Is intonation enough for TRP projection?
- How is the use of intonation integrated with other sources of information?
- What do we know about the time course of TRP projection?

Introduction: Questions

Questions adressed in this talk:

- Is intonation enough for TRP projection?
- How is the use of intonation integrated with other sources of information?
- What do we know about the time course of TRP projection?

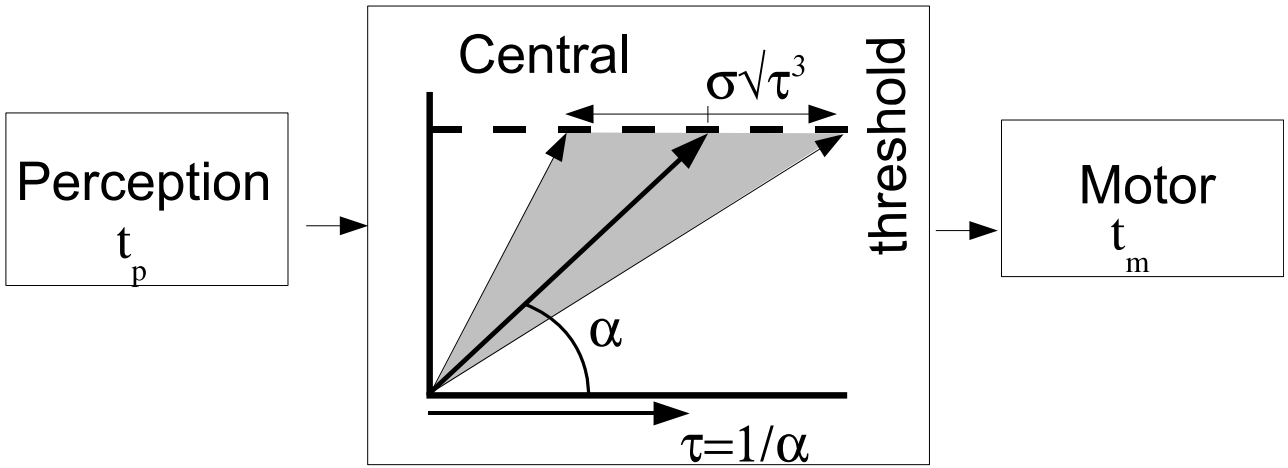
Introduction: Questions

Questions adressed in this talk:

- Is intonation enough for TRP projection?
- How is the use of intonation integrated with other sources of information?
- What do we know about the time course of TRP projection?

Introduction: Reaction-Time Model

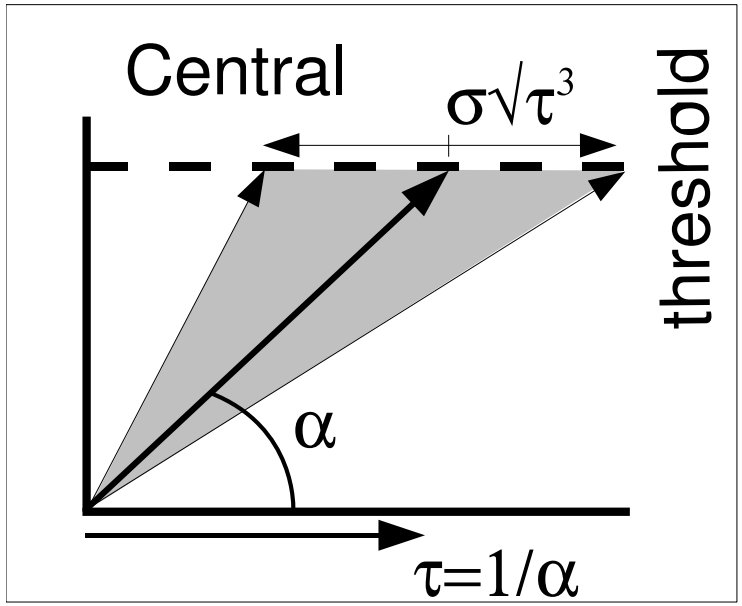
Sigman & Dehaene (2005)



Three temporal stages in Reactions to Stimuli:

- Perceptual component (P) and Motor component (M), both with deterministic response-times (t_p and t_m)
- Central **decision making component** (C) characterized by a random walk to a decision threshold
- Mean Reaction Time: $\overline{RT} = t_0 + \tau$

Introduction: Timing in PCM-model



Relative integration time to decision, τ, can be determined from the relative **variances** of the Reaction Times

$$\frac{\tau_1}{\tau_2} = \sqrt[3]{\frac{S_1^2}{S_2^2}}$$

with (S^2 = variance)

Experiment: Materials

Full Set

- 61 dialogues from CGN, telephone & face-to-face
- informal and spontaneous
- orthography, hand aligned on word level
- extra transcription on turn switches and minimal responses

Stimulus Set

- 7 telephone & 11 face-to-face dialogues (165 minutes)
- for each utterance: boundary tones are estimated as

$$Z_i > 0.2 \longrightarrow \text{high boundary tone}$$
$$-0.5 \leq Z_i \leq 0.2 \longrightarrow \text{mid boundary tone}$$
$$Z_i < -0.5 \longrightarrow \text{low boundary tone} \qquad (Z_i = \frac{\overline{F}_0 - F_0}{sd(F_0)})$$

Experiment: Materials

Full Set

- 61 dialogues from CGN, telephone & face-to-face
- informal and spontaneous
- orthography, hand aligned on word level
- extra transcription on turn switches and minimal responses

Stimulus Set

- 7 telephone & 11 face-to-face dialogues (165 minutes)
- for each utterance: boundary tones are estimated as

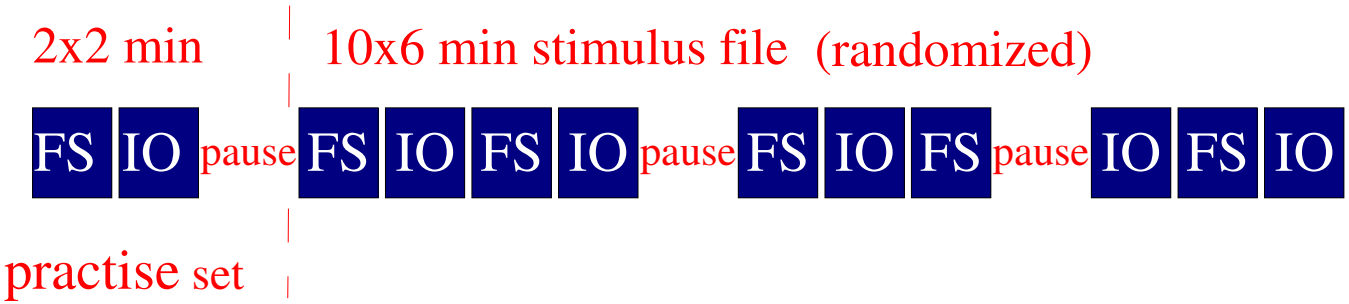
$$
\begin{aligned}
Z_i > 0.2 &\longrightarrow \text{high boundary tone} \\
-0.5 \leq Z_i \leq 0.2 &\longrightarrow \text{mid boundary tone} \\
Z_i < -0.5 &\longrightarrow \text{low boundary tone} \qquad (Z_i = \frac{\overline{F}_0 - F_0}{sd(F_0)})
\end{aligned}
$$

Experiment: Stimuli

Two sets of stimulus files:

1. **FS** Full Speech
2. **IO** Intonation Only: nothing but intonation and pause structure

   resynthesized as reiterated 'UH' sequences with the original pitch contour



Experiment: Recording Setup

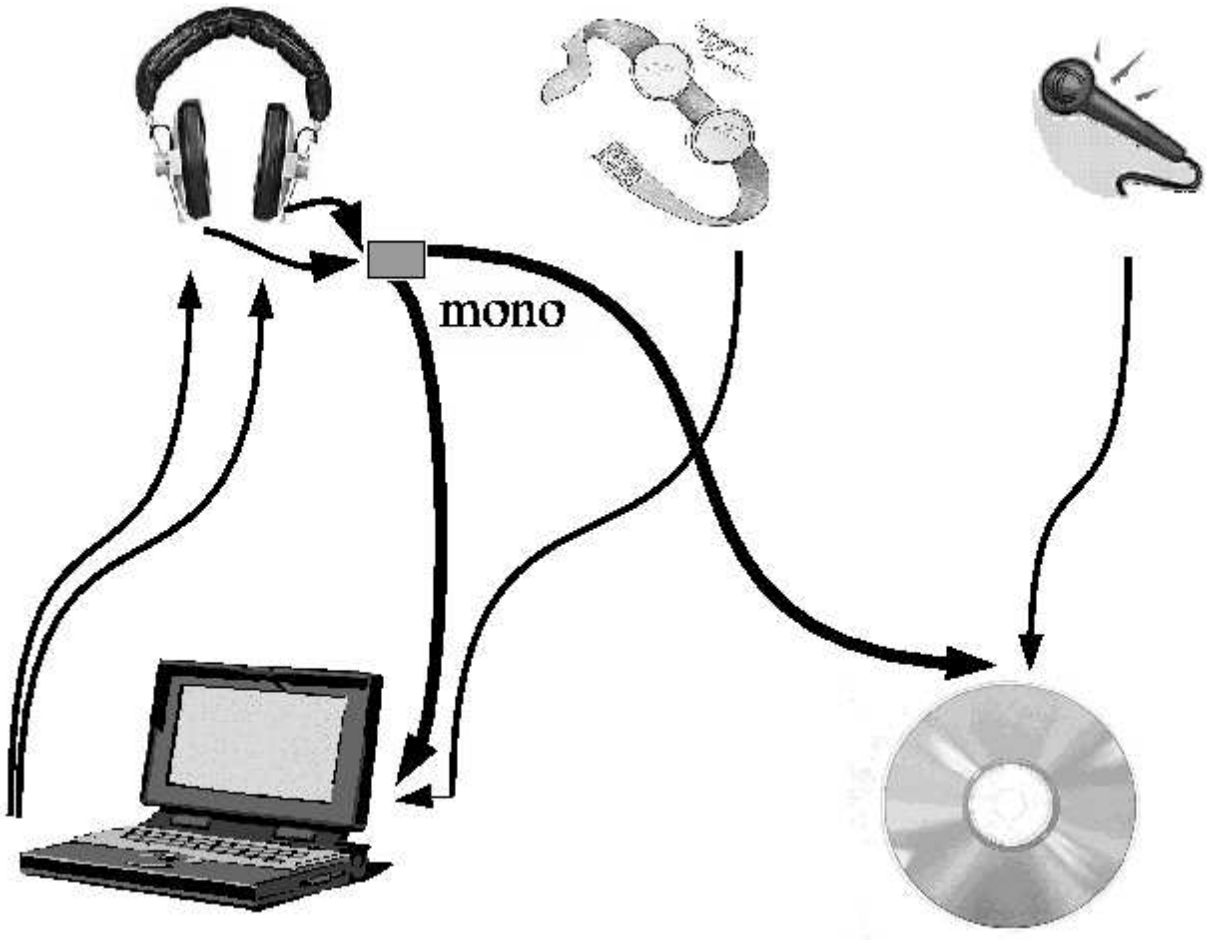
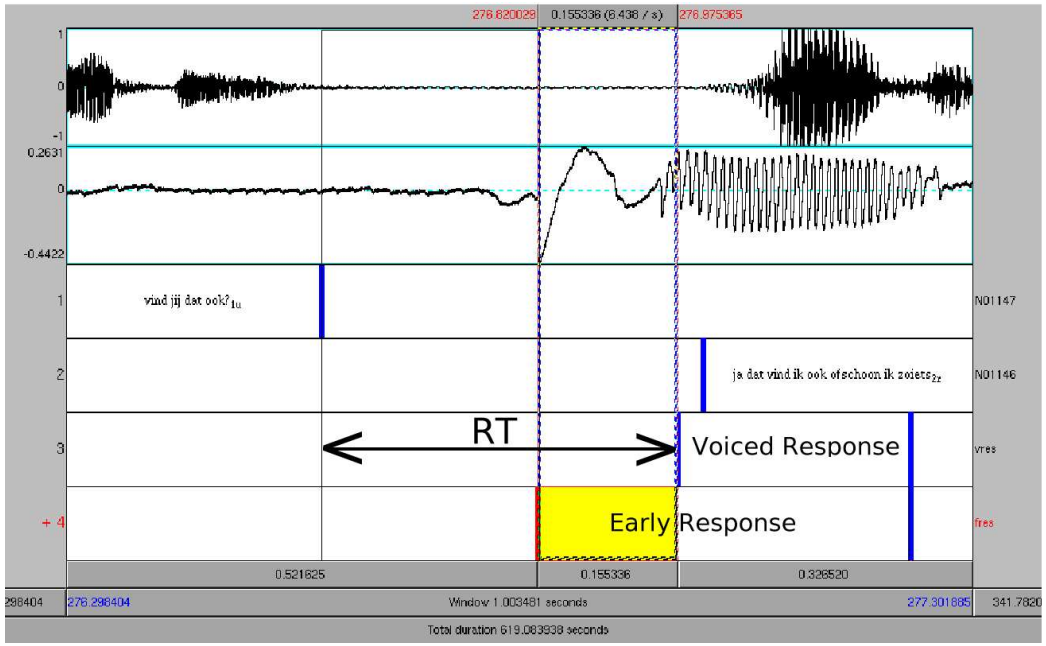


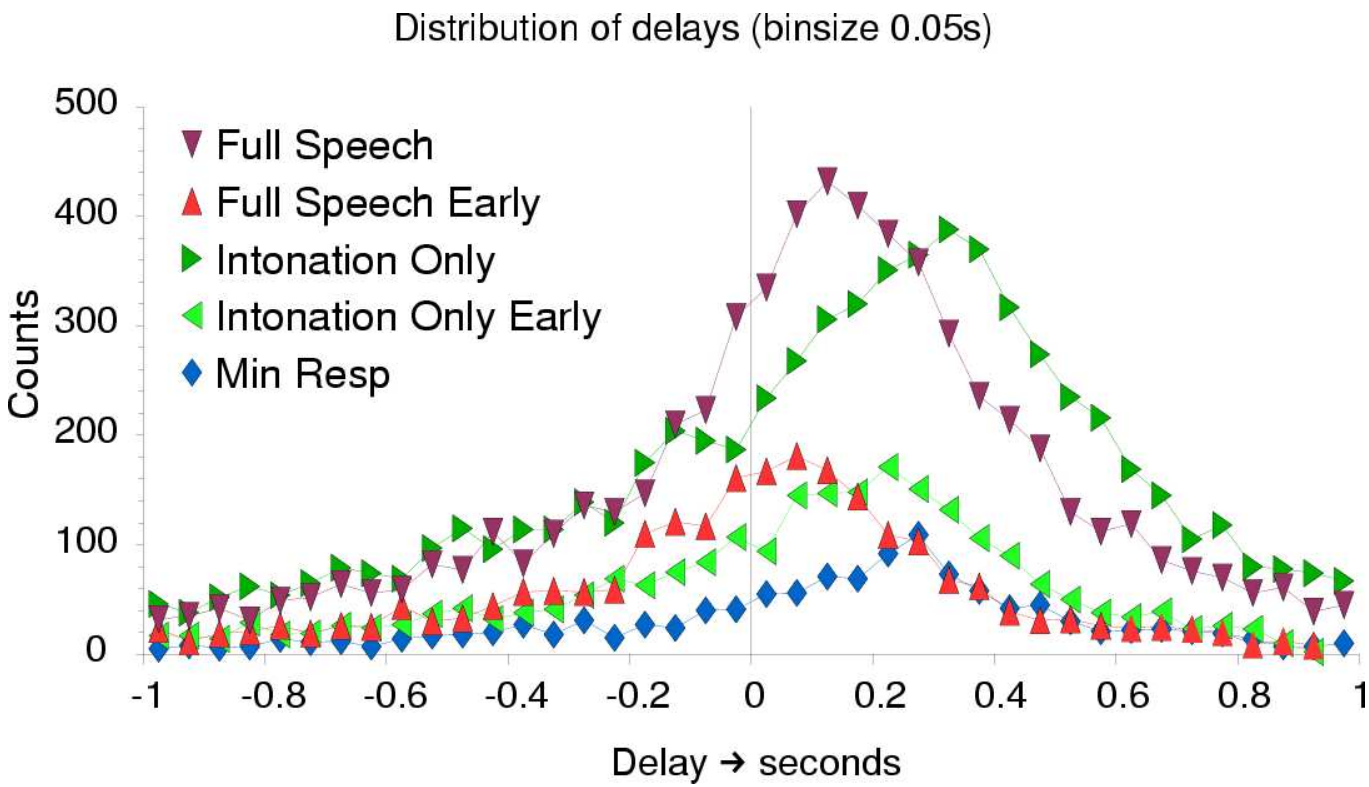
Figure: Response recording from laryngograph and microphone

Experiment: Recordings
Example response waveform and segmentation



- Top: Mono waveform of the stimulus
- Center: Laryngograph signal of a single response
- Bottom: Annotation tiers for the two speakers and the automatic segmentation of a *voiced* and *early response*.
- Intervals: The two classes of response delays and their difference in color
- Number of responses: FS/IO 6084/6575 (Early: 2349/2377)

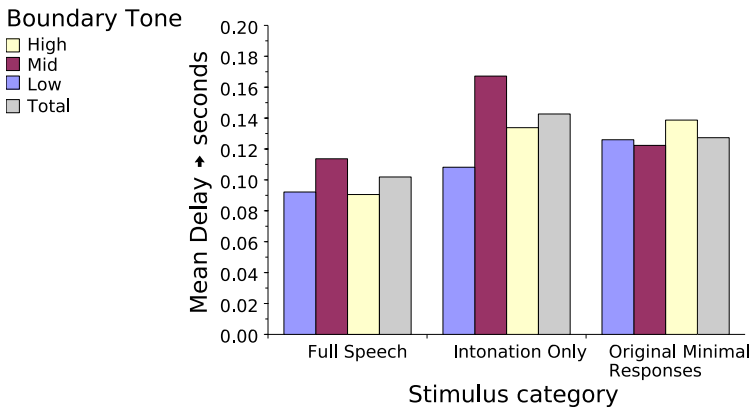
Results: Distribution of Reaction-Time Delays

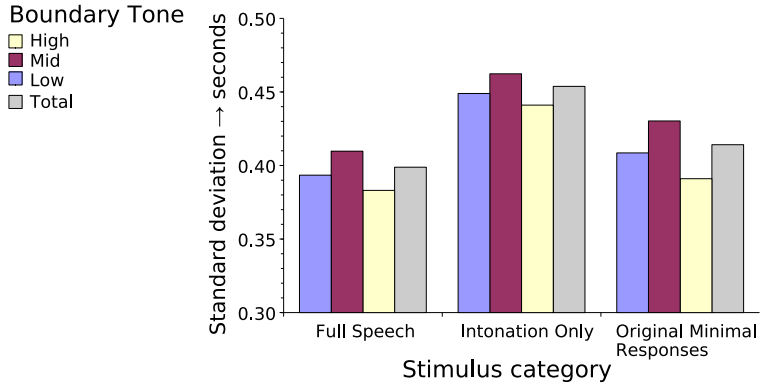


- Response counts are already increasing before end of utterance $\longrightarrow$ Projection of TRPs takes place.
- Delays are shorter for *Full Speech* stimuli (But note similar shape!)

Results: Boundary Tones

Mean Delays & Standard Deviations for Three Categories of Boundary Tones.



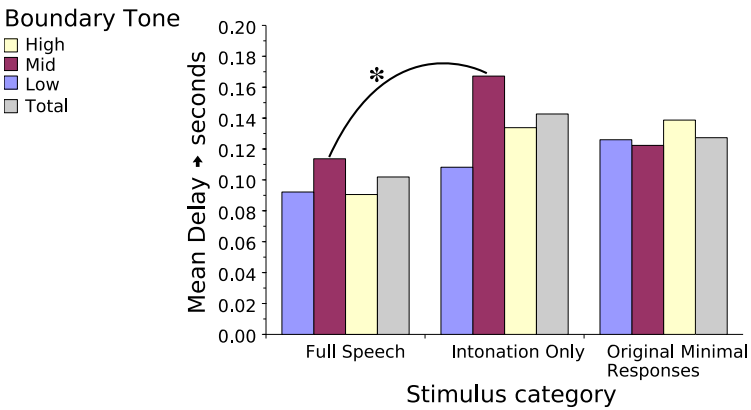


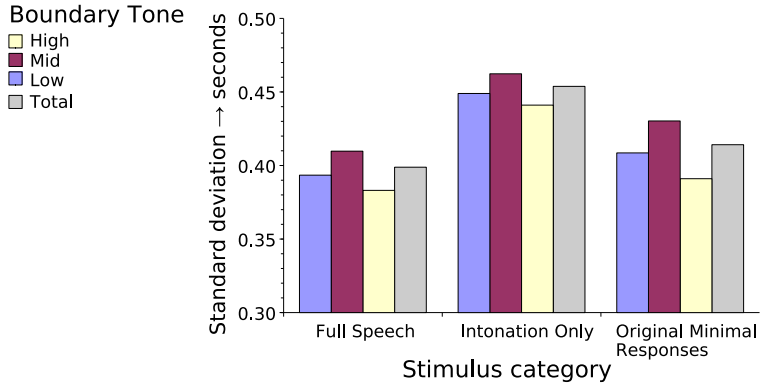
- *Intonation Only* stimuli get longer delays for mid tone endings.
- in *Intonation Only* stimuli, mid tone endings have longer delays than low and high tone endings.
- For all boundaries tones, more variance for *Intonation Only* responses
- No differences between boundary tones

*: $p < 0.01$

Results: Boundary Tones

Mean Delays & Standard Deviations for Three Categories of Boundary Tones.



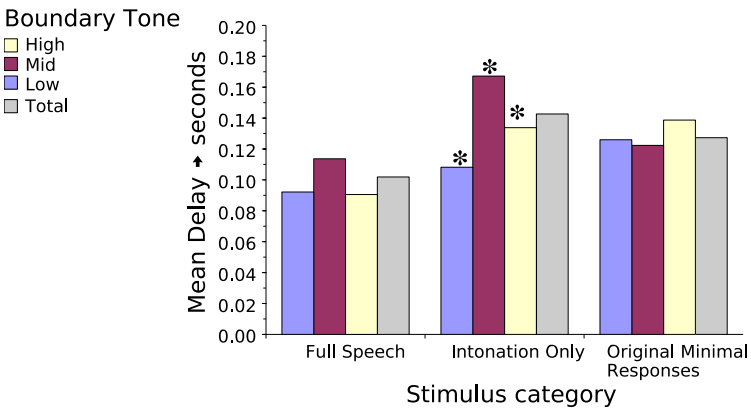


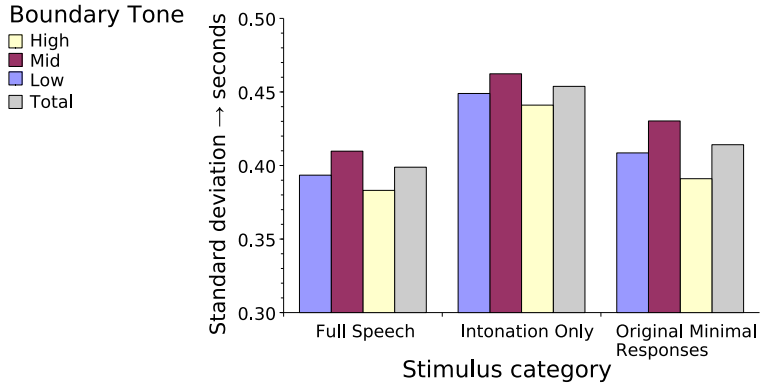
- *Intonation Only* stimuli get longer delays for mid tone endings.
- in *Intonation Only* stimuli, mid tone endings have longer delays than low and high tone endings.
- For all boundaries tones, more variance for *Intonation Only* responses
- No differences between boundary tones

*: $p < 0.01$

Results: Boundary Tones

Mean Delays & Standard Deviations for Three Categories of Boundary Tones.



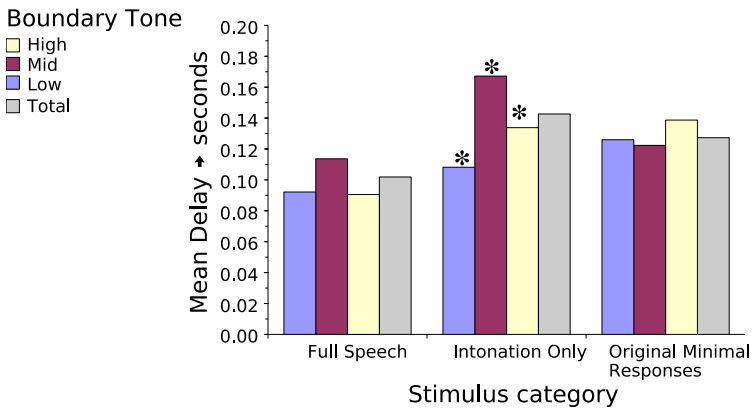


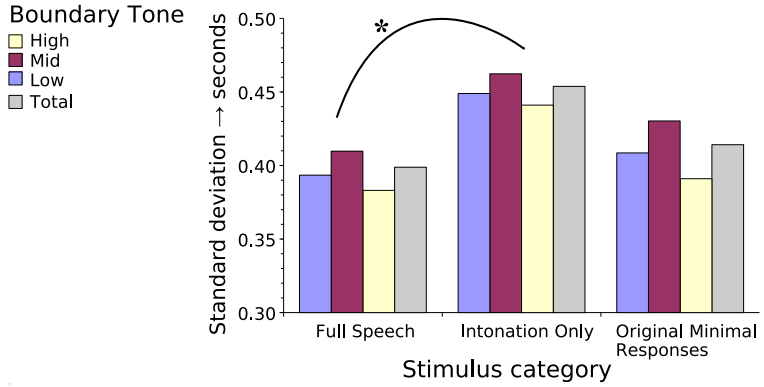
- *Intonation Only* stimuli get longer delays for mid tone endings.
- in *Intonation Only* stimuli, mid tone endings have longer delays than low and high tone endings.
- For all boundaries tones, more variance for *Intonation Only* responses
- No differences between boundary tones

*: $p < 0.01$

Results: Boundary Tones

Mean Delays & Standard Deviations for Three Categories of Boundary Tones.



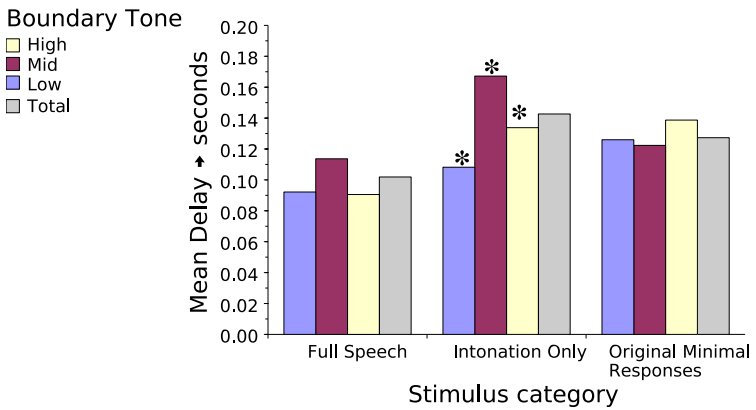


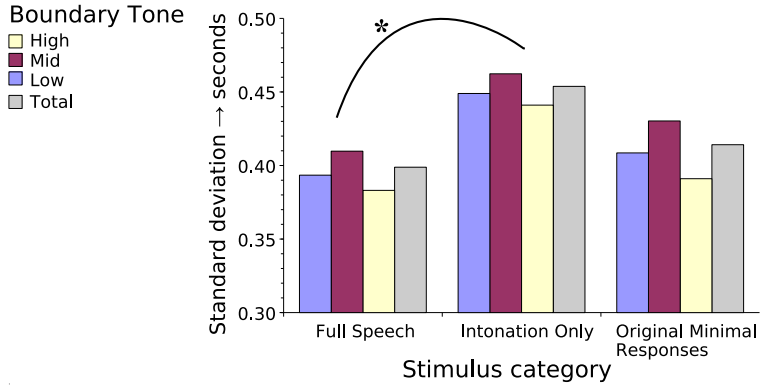
- *Intonation Only* stimuli get longer delays for mid tone endings.
- in *Intonation Only* stimuli, mid tone endings have longer delays than low and high tone endings.
- For all boundaries tones, more variance for *Intonation Only* responses
- No differences between boundary tones

*: $p < 0.01$

Results: Boundary Tones

Mean Delays & Standard Deviations for Three Categories of Boundary Tones.



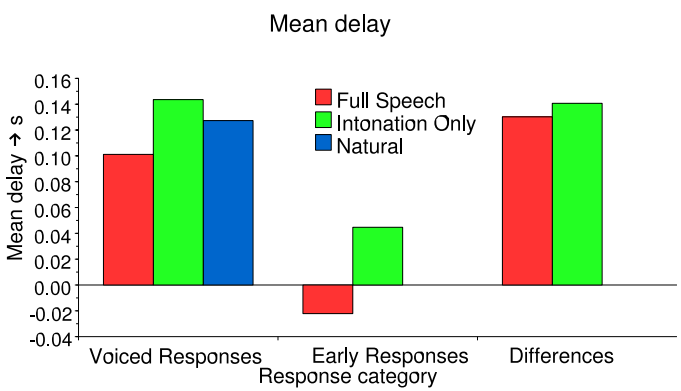


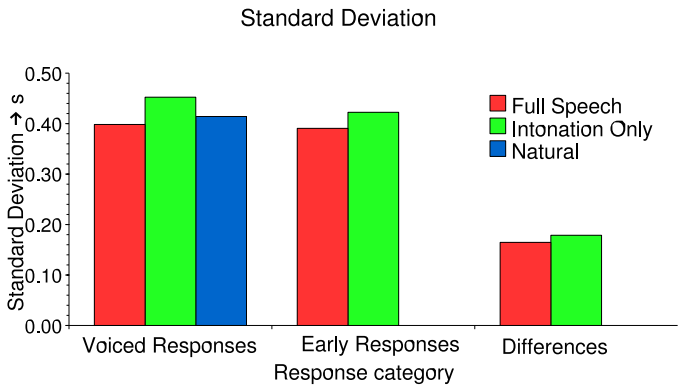
- *Intonation Only* stimuli get longer delays for mid tone endings.
- in *Intonation Only* stimuli, mid tone endings have longer delays than low and high tone endings.
- For all boundaries tones, more variance for *Intonation Only* responses
- No differences between boundary tones

*: $p < 0.01$

Results: Early Responses

Mean Delays & Standard Deviations for Three Types of Response Delays.

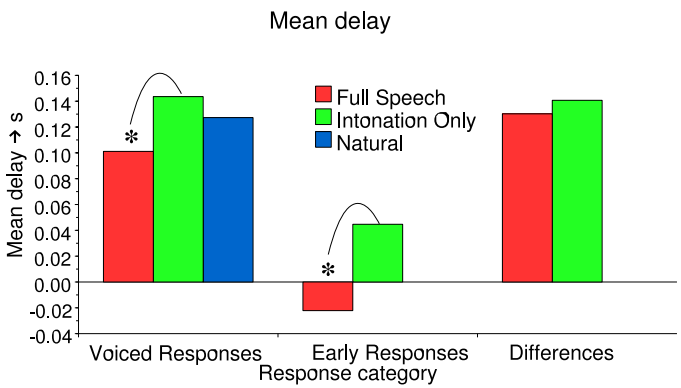


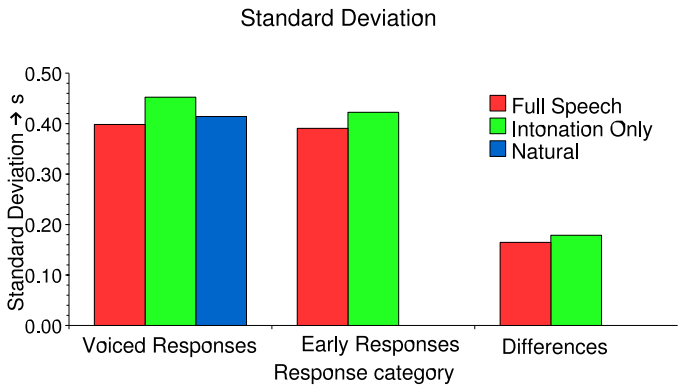


- NB: *Early* & *voiced* resp. differ by construction!
- Mean delays for *FS* are shorter than those for *IO* for both *voiced* and *early responses*.
- The mean delay of the difference RT is also longer for *IO* stimuli.
- More variance in responses to *IO* stimuli for both *voiced* and *early responses*.
- No difference in variance of the difference RTs.
- The variance of the difference Rts was much lower than the variance of the *voiced* and *early* RTs.

Results: Early Responses

Mean Delays & Standard Deviations for Three Types of Response Delays.

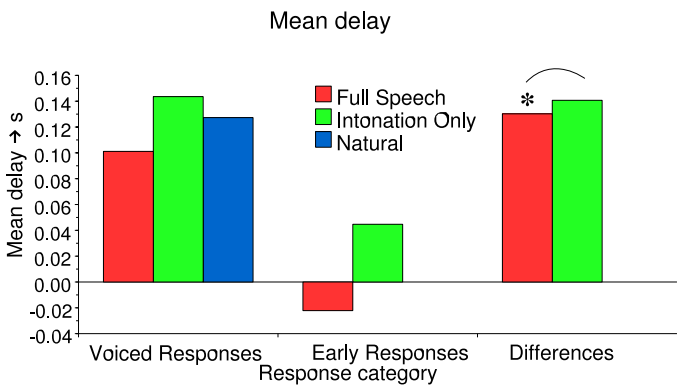


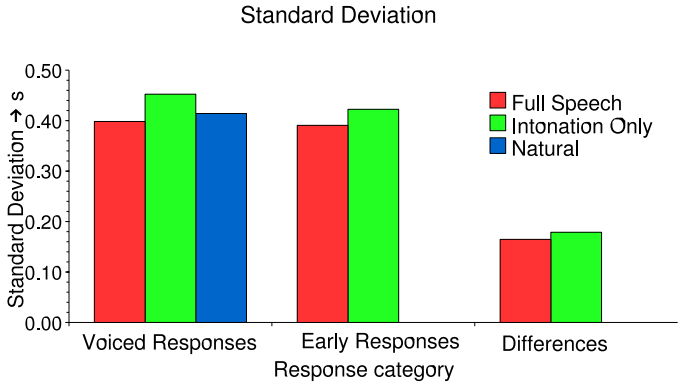


- NB: *Early* & *voiced* resp. differ by construction!
- Mean delays for *FS* are shorter than those for *IO* for both *voiced* and *early responses*.
- The mean delay of the difference RT is also longer for *IO* stimuli.
- More variance in responses to *IO* stimuli for both *voiced* and *early responses*.
- No difference in variance of the difference RTs.
- The variance of the difference Rts was much lower than the variance of the *voiced* and *early* RTs.

Results: Early Responses

Mean Delays & Standard Deviations for Three Types of Response Delays.

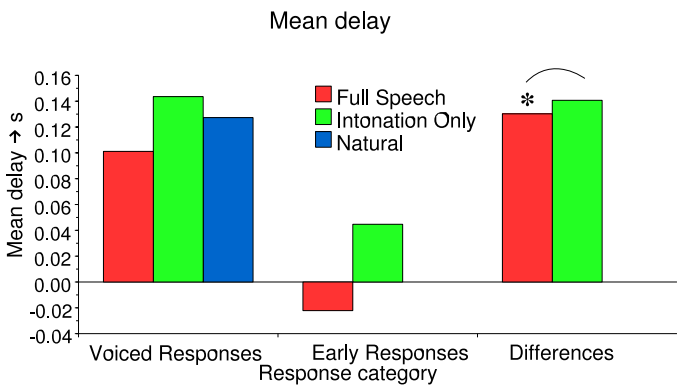


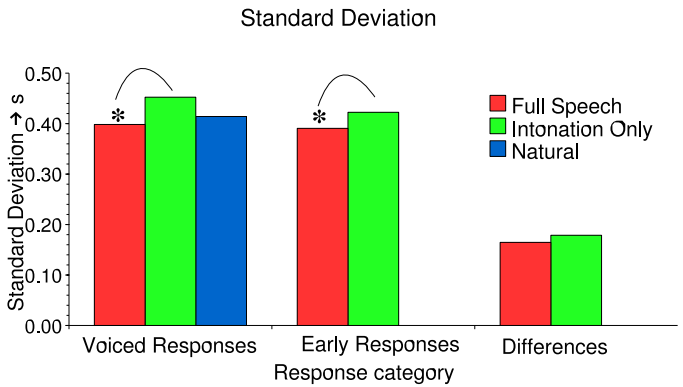


- NB: *Early* & *voiced* resp. differ by construction!
- Mean delays for *FS* are shorter than those for *IO* for both *voiced* and *early responses*.
- The mean delay of the difference RT is also longer for *IO* stimuli.
- More variance in responses to *IO* stimuli for both *voiced* and *early responses*.
- No difference in variance of the difference RTs.
- The variance of the difference Rts was much lower than the variance of the *voiced* and *early* RTs.

Results: Early Responses

Mean Delays & Standard Deviations for Three Types of Response Delays.



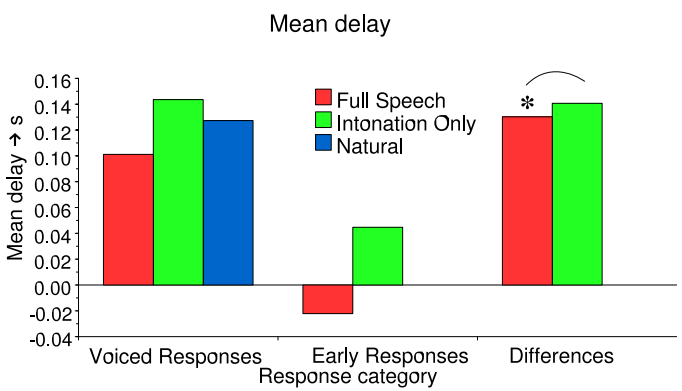


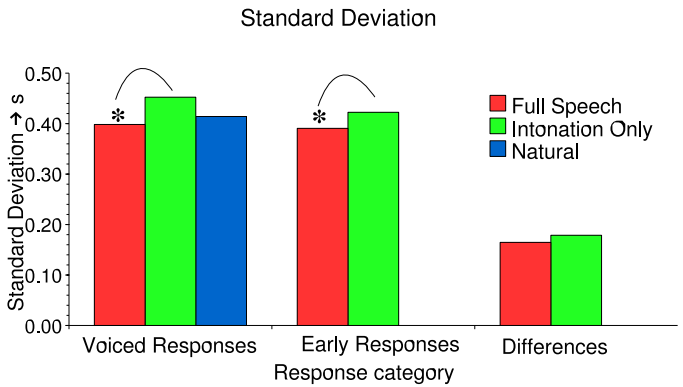
- NB: *Early* & *voiced* resp. differ by construction!
- Mean delays for *FS* are shorter than those for *IO* for both *voiced* and *early responses*.
- The mean delay of the difference RT is also longer for *IO* stimuli.

- More variance in responses to *IO* stimuli for both *voiced* and *early responses*.
- No difference in variance of the difference RTs.
- The variance of the difference Rts was much lower than the variance of the *voiced* and *early* RTs.

Results: Early Responses

Mean Delays & Standard Deviations for Three Types of Response Delays.



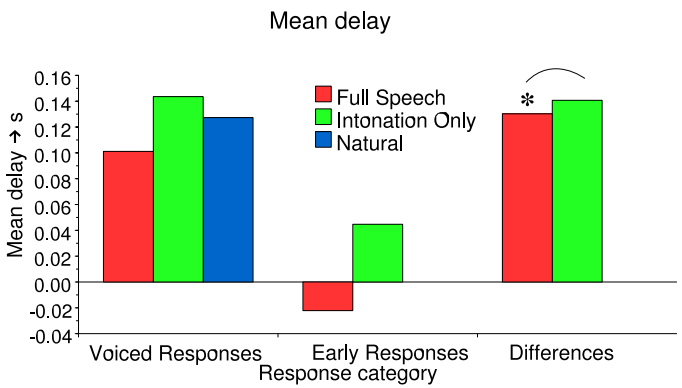


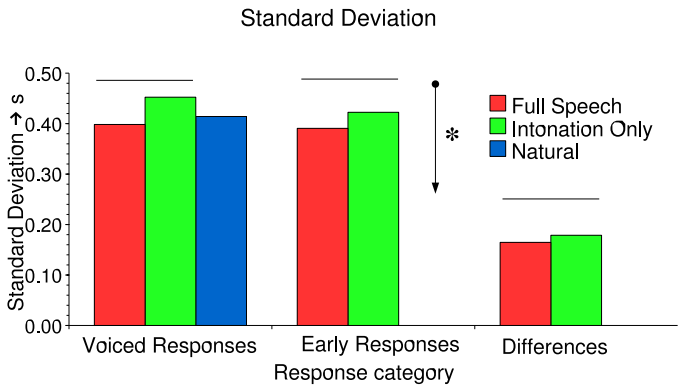
- NB: *Early* & *voiced* resp. differ by construction!
- Mean delays for *FS* are shorter than those for *IO* for both *voiced* and *early responses*.
- The mean delay of the difference RT is also longer for *IO* stimuli.

- More variance in responses to *IO* stimuli for both *voiced* and *early responses*.
- No difference in variance of the difference RTs.
- The variance of the difference Rts was much lower than the variance of the *voiced* and *early* RTs.

Results: Early Responses

Mean Delays & Standard Deviations for Three Types of Response Delays.





- NB: *Early* & *voiced* resp. differ by construction!
- Mean delays for *FS* are shorter than those for *IO* for both *voiced* and *early responses*.
- The mean delay of the difference RT is also longer for *IO* stimuli.

- More variance in responses to *IO* stimuli for both *voiced* and *early responses*.
- No difference in variance of the difference RTs.
- The variance of the difference Rts was much lower than the variance of the *voiced* and *early* RTs.

Discussion: Effect of Boundary tones

First question:

- Is intonation enough for TRP projection?

Discussion: Effect of Boundary tones

First question:

- Is intonation enough for TRP projection?

- *Intonation Only* responses are delayed for *mid tone* endings) & they have more variance.

Discussion: Effect of Boundary tones

First question:

- Is intonation enough for TRP projection?

- *Intonation Only* responses are delayed for *mid tone* endings) & they have more variance.
- Still faster than most latencies for shadowing tasks

Discussion: Effect of Boundary tones

First question:

- Is intonation enough for TRP projection?

- *Intonation Only* responses are delayed for *mid tone* endings) & they have more variance.
- Still faster than most latencies for shadowing tasks
- Rapid responses + effect of boundary tones rule out that subjects reacted to the utterance ends themselves.

Discussion: Effect of Boundary tones

First question:

- Is intonation enough for TRP projection?

- *Intonation Only* responses are delayed for *mid tone* endings) & they have more variance.
- Still faster than most latencies for shadowing tasks
- Rapid responses + effect of boundary tones rule out that subjects reacted to the utterance ends themselves.

  ➜ Mid tones: subjects have to wait for the pause.

Discussion: Effect of Boundary tones

First question:

- Is intonation enough for TRP projection?

- *Intonation Only* responses are delayed for *mid tone* endings) & they have more variance.
- Still faster than most latencies for shadowing tasks
- Rapid responses + effect of boundary tones rule out that subjects reacted to the utterance ends themselves.

➜ Mid tones: subjects have to wait for the pause.
➜ Intonation into a high or low boundary tone is sufficient to predict an upcoming utterance end,
at least some of the time.

Discussion: Integration of Intonation

Second question:

- How is the use of intonation integrated with other sources of information?

- Both boundary tones and verbal and prosodic information help TRP projection (reduced delays)
- The difference between *voiced* and *early responses* was not affected by the stimulus-type
- *Intonation Only* stimuli mostly affect *early* integration-times, not the timing after *early responses*.

➔ There seems to be a perceptual, P, type of delay.

➔ Removing everything but intonation & pauses increases the integration time with around 10 ± 1.3 %

Discussion: Integration of Intonation

Second question:

- How is the use of intonation integrated with other sources of information?

- Both boundary tones and verbal and prosodic information help TRP projection (reduced delays)
- The difference between *voiced* and *early responses* was not affected by the stimulus-type
- *Intonation Only* stimuli mostly affect *early* integration-times, not the timing after *early responses*.

➔ There seems to be a perceptual, P, type of delay.

➔ Removing everything but intonation & pauses increases the integration time with around 10 ± 1.3 %

Discussion: Integration of Intonation

Second question:

- How is the use of intonation integrated with other sources of information?

- Both boundary tones and verbal and prosodic information help TRP projection (reduced delays)
- The difference between *voiced* and *early responses* was not affected by the stimulus-type
- *Intonation Only* stimuli mostly affect *early* integration-times, not the timing after *early responses*.

➔ There seems to be a perceptual, P_i, type of delay.

➔ Removing everything but intonation & pauses increases the integration time with around 10 ± 1.3 %

Discussion: Integration of Intonation

Second question:

- How is the use of intonation integrated with other sources of information?

- Both boundary tones and verbal and prosodic information help TRP projection (reduced delays)
- The difference between *voiced* and *early responses* was not affected by the stimulus-type
- *Intonation Only* stimuli mostly affect *early* integration-times, not the timing after *early responses*.

➜ There seems to be a perceptual, P, type of delay.
➜ Removing everything but intonation & pauses increases the integration time with around 10 ± 1.3 %

Discussion: Integration of Intonation

Second question:

- How is the use of intonation integrated with other sources of information?

- Both boundary tones and verbal and prosodic information help TRP projection (reduced delays)
- The difference between *voiced* and *early responses* was not affected by the stimulus-type
- *Intonation Only* stimuli mostly affect *early* integration-times, not the timing after *early responses*.

➜ There seems to be a perceptual, P, type of delay.

➜ Removing everything but intonation & pauses increases the integration time with around 10 ± 1.3 %

Discussion: Integration of Intonation

Second question:

- How is the use of intonation integrated with other sources of information?

- Both boundary tones and verbal and prosodic information help TRP projection (reduced delays)
- The difference between *voiced* and *early responses* was not affected by the stimulus-type
- *Intonation Only* stimuli mostly affect *early* integration-times, not the timing after *early responses*.

➜ There seems to be a perceptual, P, type of delay.
➜ Removing everything but intonation & pauses increases the integration time with around 10 ± 1.3 %

Discussion: Time Course of TRP Projection

Third question:

- What do we know about the time course of TRP projection?

- We can determine the relative amounts of (integration) time for early and voiced responses $\frac{\tau_{diff}}{\tau_{early}} \approx 0.55$
- Early integration time τ_{early} is about 2 x difference integration time τ_{diff}
- $\tau_{voiced} = \tau_{early} + \tau_{diff} \Leftrightarrow \tau_{diff} = RT_{voiced} - RT_{early}$

  ➔ With a t_0 of $\geq$50 ms under the most favorable circumstances (shadowing tasks) we can conclude that planning (elicited) minimal responses starts more than 300 ms before the actual utterance end (TRP).

Discussion: Time Course of TRP Projection

Third question:

- What do we know about the time course of TRP projection?

- We can determine the relative amounts of (integration) time for early and voiced responses $\frac{\tau_{diff}}{\tau_{early}} \approx 0.55$
- Early integration time τ_{early} is about 2 x difference integration time τ_{diff}
- $\tau_{voiced} = \tau_{early} + \tau_{diff} \Leftrightarrow \tau_{diff} = RT_{voiced} - RT_{early}$

  ➜ With a t_0 of $\geq$50 ms under the most favorable circumstances (shadowing tasks) we can conclude that planning (elicited) minimal responses starts more than 300 ms before the actual utterance end (TRP).

Discussion: Time Course of TRP Projection

Third question:

- What do we know about the time course of TRP projection?

- We can determine the relative amounts of (integration) time for early and voiced responses $\frac{\tau_{diff}}{\tau_{early}} \approx 0.55$
- Early integration time τ_{early} is about 2 x difference integration time τ_{diff}
- $\tau_{voiced} = \tau_{early} + \tau_{diff} \Leftrightarrow \tau_{diff} = RT_{voiced} - RT_{early}$

  ➔ With a t_0 of $\geq$50 ms under the most favorable circumstances (shadowing tasks) we can conclude that planning (elicited) minimal responses starts more than 300 ms before the actual utterance end (TRP).

Discussion: Time Course of TRP Projection

Third question:

- What do we know about the time course of TRP projection?

- We can determine the relative amounts of (integration) time for early and voiced responses $\frac{\tau_{diff}}{\tau_{early}} \approx 0.55$
- Early integration time τ_{early} is about 2 x difference integration time τ_{diff}
- $\tau_{voiced} = \tau_{early} + \tau_{diff} \Leftrightarrow \tau_{diff} = RT_{voiced} - RT_{early}$

  ➜ With a t_0 of $\geq$50 ms under the most favorable circumstances (shadowing tasks) we can conclude that planning (elicited) minimal responses starts more than 300 ms before the actual utterance end (TRP).

Discussion: Time Course of TRP Projection

Third question:

- What do we know about the time course of TRP projection?

- We can determine the relative amounts of (integration) time for early and voiced responses $\frac{\tau_{diff}}{\tau_{early}} \approx 0.55$
- Early integration time τ_{early} is about 2 x difference integration time τ_{diff}
- $\tau_{voiced} = \tau_{early} + \tau_{diff} \Leftrightarrow \tau_{diff} = RT_{voiced} - RT_{early}$

  ➜ With a t_0 of $\geq$50 ms under the most favorable circumstances (shadowing tasks) we can conclude that planning (elicited) minimal responses starts more than 300 ms before the actual utterance end (TRP).

Conclusions

- End-intonation sufficient cue for upcoming TRP in *intonation only* stimuli, but: more time is needed to predict an utterance end
- Subjects can predict an upcoming TRP from *high* or *low* boundary tones
- but, most likely, have to wait until they perceive the end of the utterance (pause) in *mid* boundary tone *intonation only* stimuli
- The articulation of elicited minimal responses has at least one intermediate stage, visible as an articulatory preparation step.
- Planning (elicited) minimal responses starts more than 300 ms before the utterance end (TRP).

Thank you!

Caspers J., "Local speech melody as a limiting factor in the turn-taking system in Dutch", Journal of Phonetics 31: 139-278, 2003.

Sigman M., Dehaene S., "Parsing a Cognitive Task: A Characterization of the Mind's Bottleneck", PLoS Biology 3, e37, 2005.

Appendix: Formulas

Probability of a random walk crossing a threshold for the first time at time t:

$$g(t) = \frac{1}{\sigma \cdot \sqrt{2\pi \cdot (t - t_0)^3}} \cdot exp\left(-\frac{(1 - \alpha \cdot (t - t_0))^2}{2 \cdot \sigma^2 (t - t_0)}\right) \quad (1)$$

Mean Reaction Time: $\overline{RT} = t_0 + \tau$

Variation of Reaction Time: $var(RT) = \frac{1}{2}\sigma^2\tau^3$

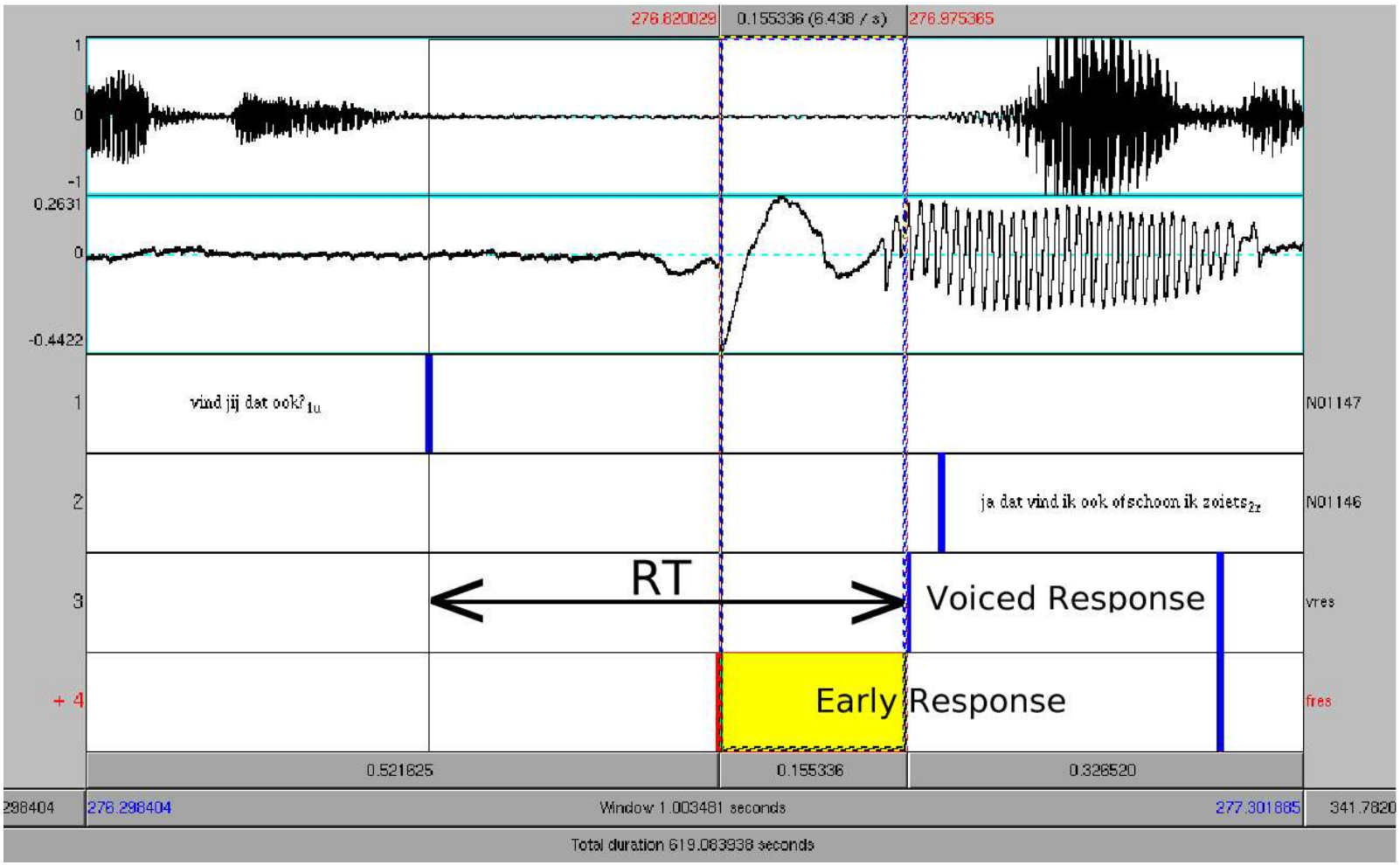
Relative Integration Times: $\frac{\tau_i}{\tau_j} = \sqrt[3]{\frac{s_i^2}{s_j^2}}$

Appendix: Calculations

- Relative amounts of (integration) time for τ_{early} and τ_{diff}, $\frac{\tau_{diff}}{\tau_{early}} \approx 0.55$
- $\longrightarrow \tau_{early}$ is about 2 x τ_{diff}
- With a simple model: $\tau_{voiced} = \tau_{early} + \tau_{diff}$
  $\Leftrightarrow \tau_{diff} = RT_{voiced} - RT_{early}$
- For *full speech*, average *difference* RT is 130 ms, integration-time, τ_{early}, is 235 ms and the total effective integration-times τ_{voiced} is 370 ms
- For *intonation only*, the average *difference* RT is 140 ms, τ_{early} is 255 ms and τ_{voiced} is 400 ms.
- With a t_0 of $\geq$50 ms (taken from shadowing tasks), planning starts more than 300 ms before the actual utterance end.

◂ Discussion

Appendix: Recordings



Appendix: Reaction Time Distribution under PCM model

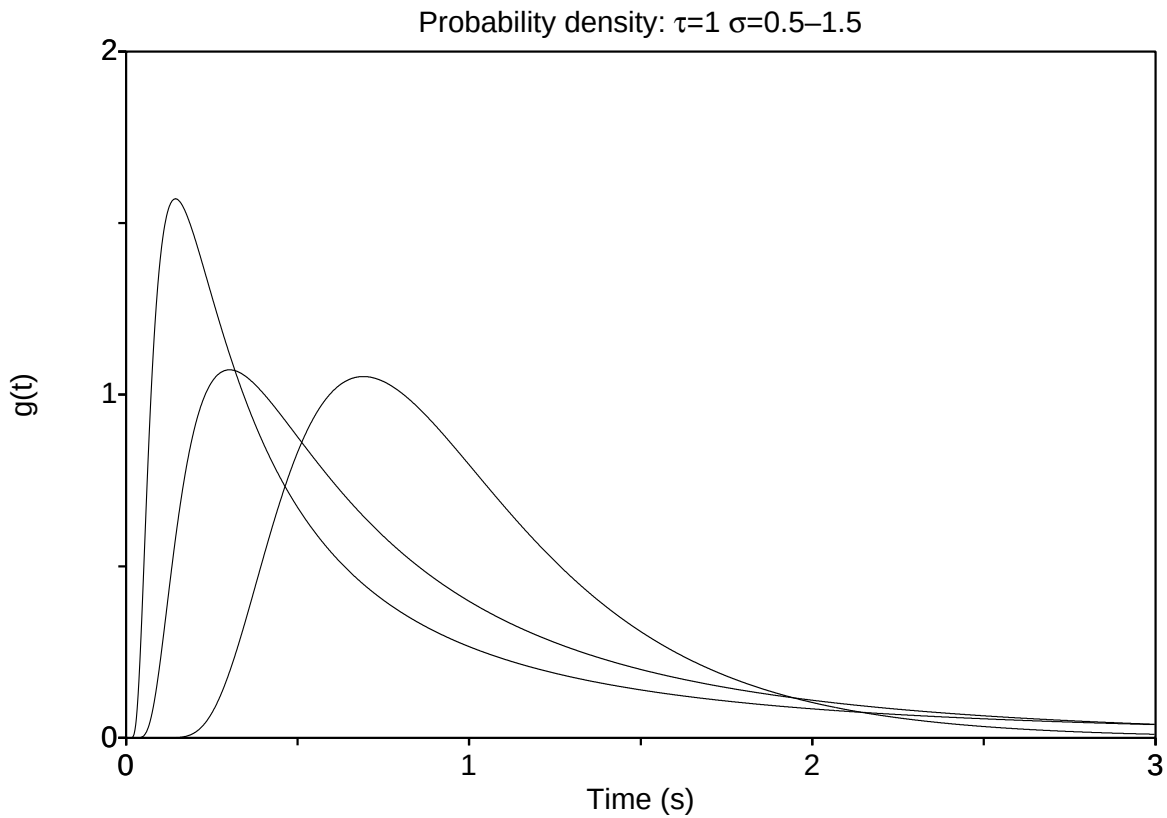


Figure: Distribution of RTs for $\tau = 1$ and $\sigma = [1.5, 1.0, 0.5]$

Number of Responses

Table: *Total number of articulated (voiced) and early responses to stimuli for each of the 3 end-tone categories and minimal responses for the total conversation set.*

response category	low	mid	high	total
full speech voiced	1860	2850	1374	6084
early	690	1144	515	2349
intonation only voiced	1917	3205	1453	6575
early	663	1180	534	2377
full dialog set (min. resp.)	386	539	281	1206

For roughly $\frac{1}{3}$ of all responses we can measure a so called *Early Response*