Prosody and the brain II: Turn-taking

Sara Bögels

Max Planck Institute for Psycholinguistics

Overview

Turn-taking models

Turn-taking & non-prosodic cues

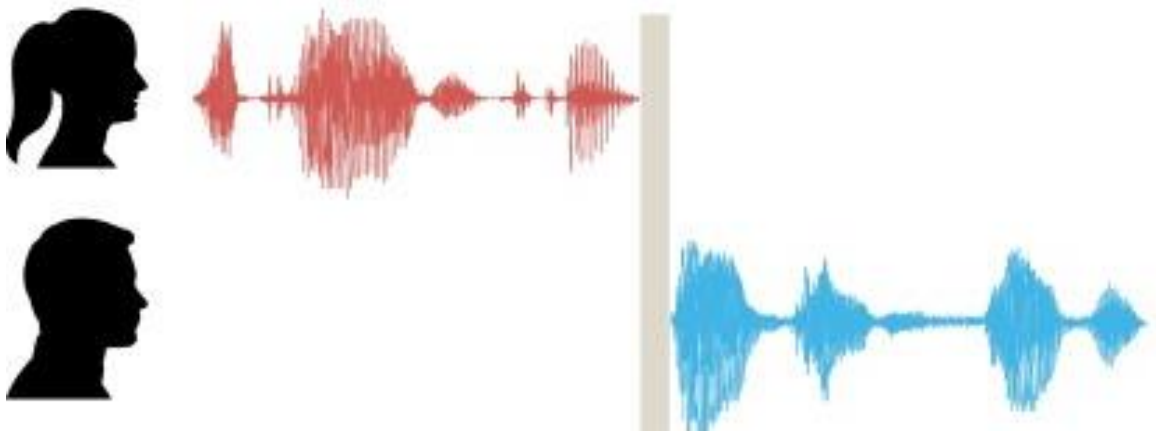
- early anticipation (EEG)
- early planning (EEG)

Turn-taking & final prosodic cues

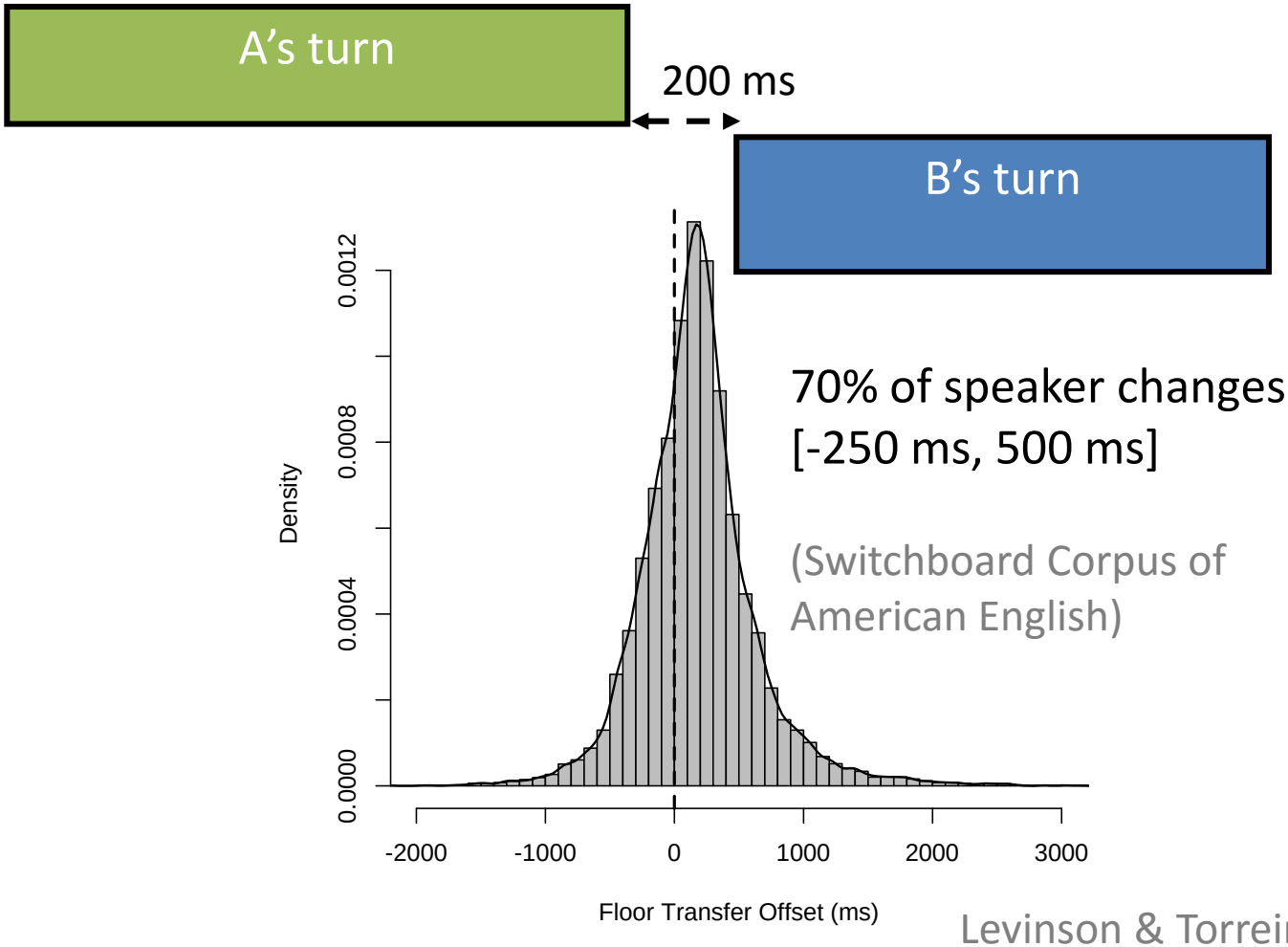
- importance of final information
- importance of prosodic cues

Turn-taking & silence (EEG)

Turn-taking models



Turn-taking

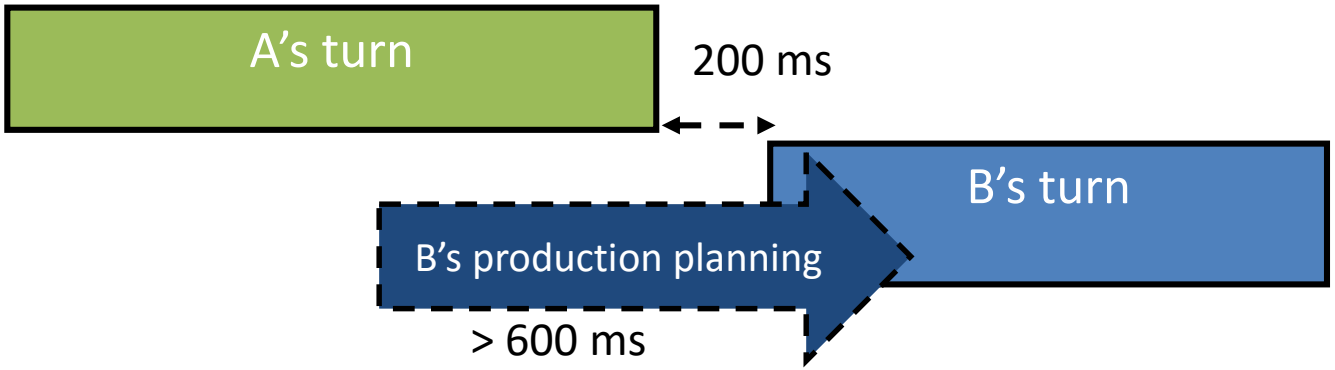


70% of speaker changes
[-250 ms, 500 ms]

(Switchboard Corpus of American English)

Levinson & Torreira (2015)

A psycholinguistic puzzle

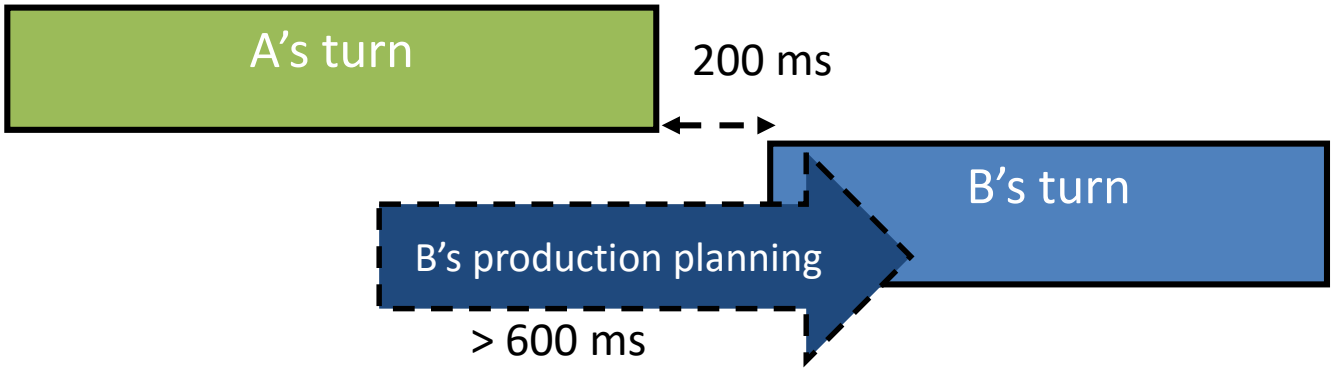


Planning and producing language takes time, e.g.:

- picture naming: 600 ms Levelt et al. (1999)
- simple sentence (SVO) production: 1500 ms Griffin & Bock (2000)

- Planning starts in overlap with the current turn
- The turn end should be estimated precisely

A psycholinguistic puzzle

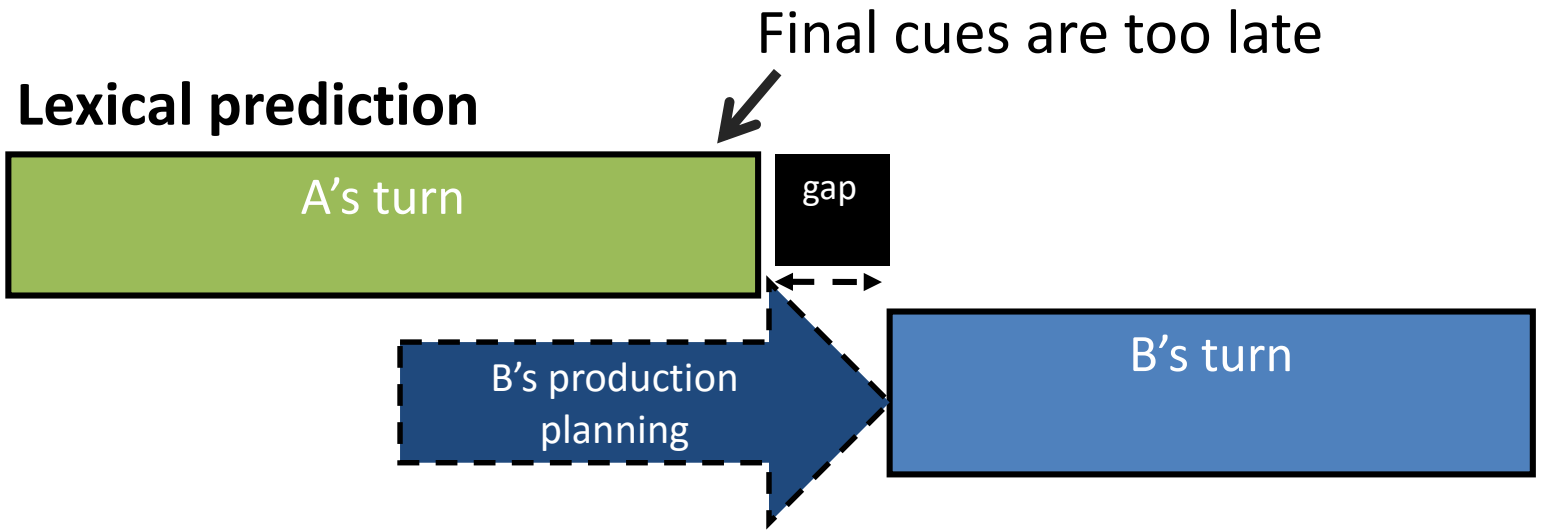


Planning and producing language takes time, e.g.:

- picture naming: 600 ms Levelt et al. (1999)
- simple sentence (SVO) production: 1500 ms Griffin & Bock (2000)

- Planning starts in overlap with the current turn
- The turn end should be estimated precisely

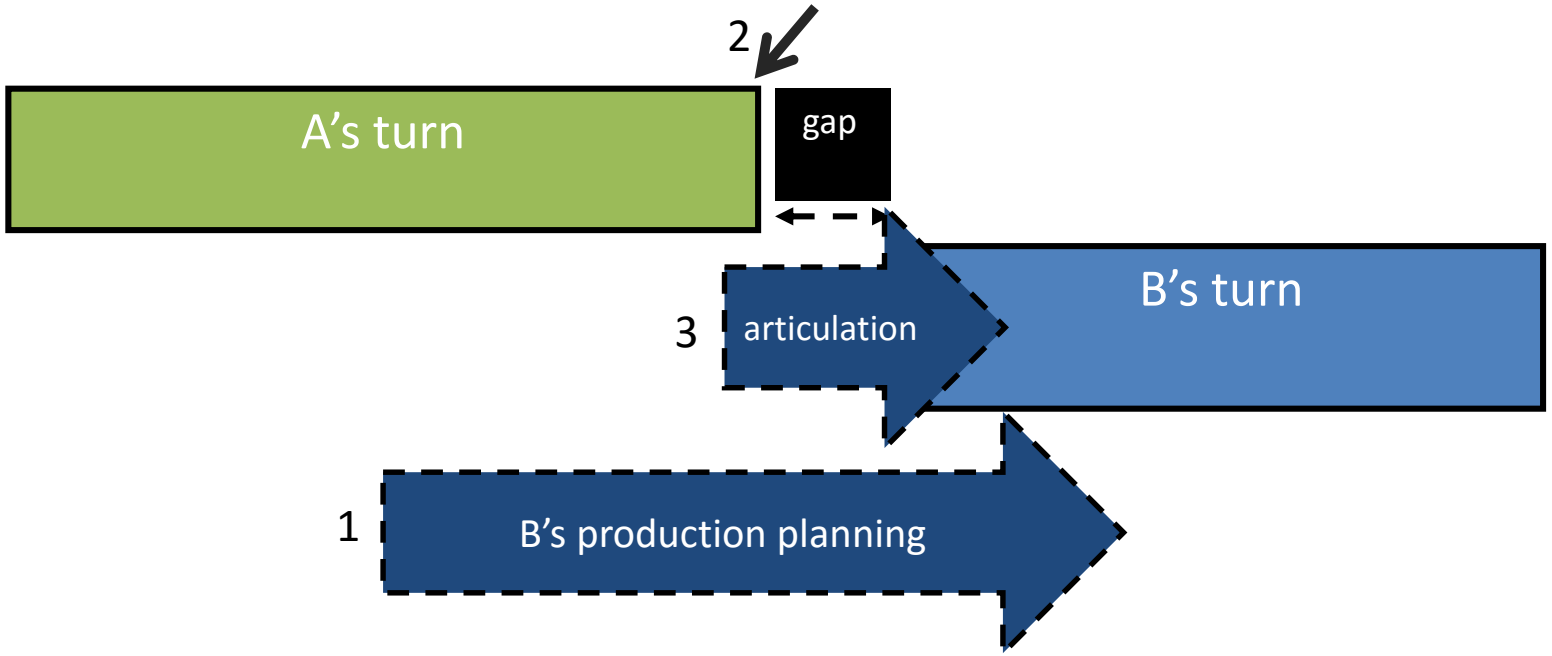
Model 1



1. Predict words up to end of turn
2. Estimate duration of these words
3. Launch planning to start speaking around estimated turn end

Sacks et al. (1974); De Ruiter, Mitterer & Enfield (2006)

Model 2

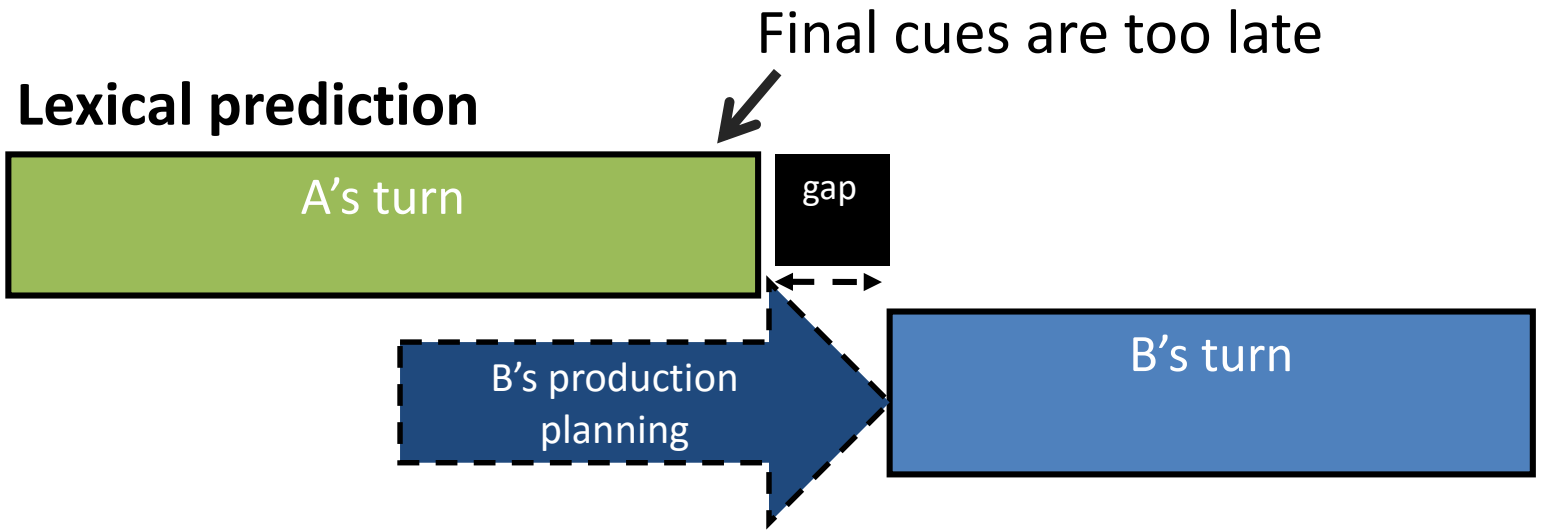


1. Early conceptual and linguistic planning
2. Identification of turn-completion
3. Launch articulation

e.g., Levinson & Torreira (2015)

Turn-taking & non-prosodic effects

Model 1: Anticipation



1. Predict words up to end of turn
2. Estimate duration of these words
3. Launch planning to start speaking around estimated turn end

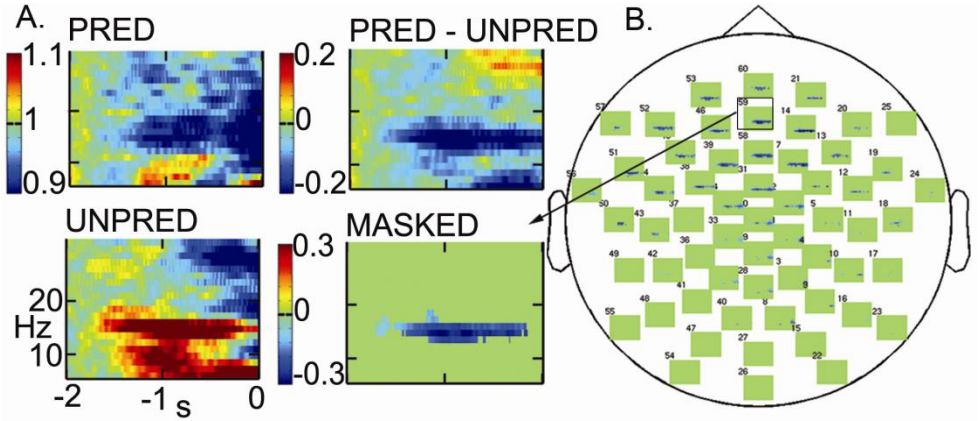
Sacks et al. (1974); De Ruiter, Mitterer & Enfield (2006)

Turn-end anticipation (1)

Task: Listen to isolated conversational turns
Press a button exactly at the moment when the turn ends

De Ruiter, Mitterer, & Enfield (2006)

Predictable & Unpredictable turns
Lower beta (11-18.5Hz) power for predictable turns
from 1.8 s before button-press



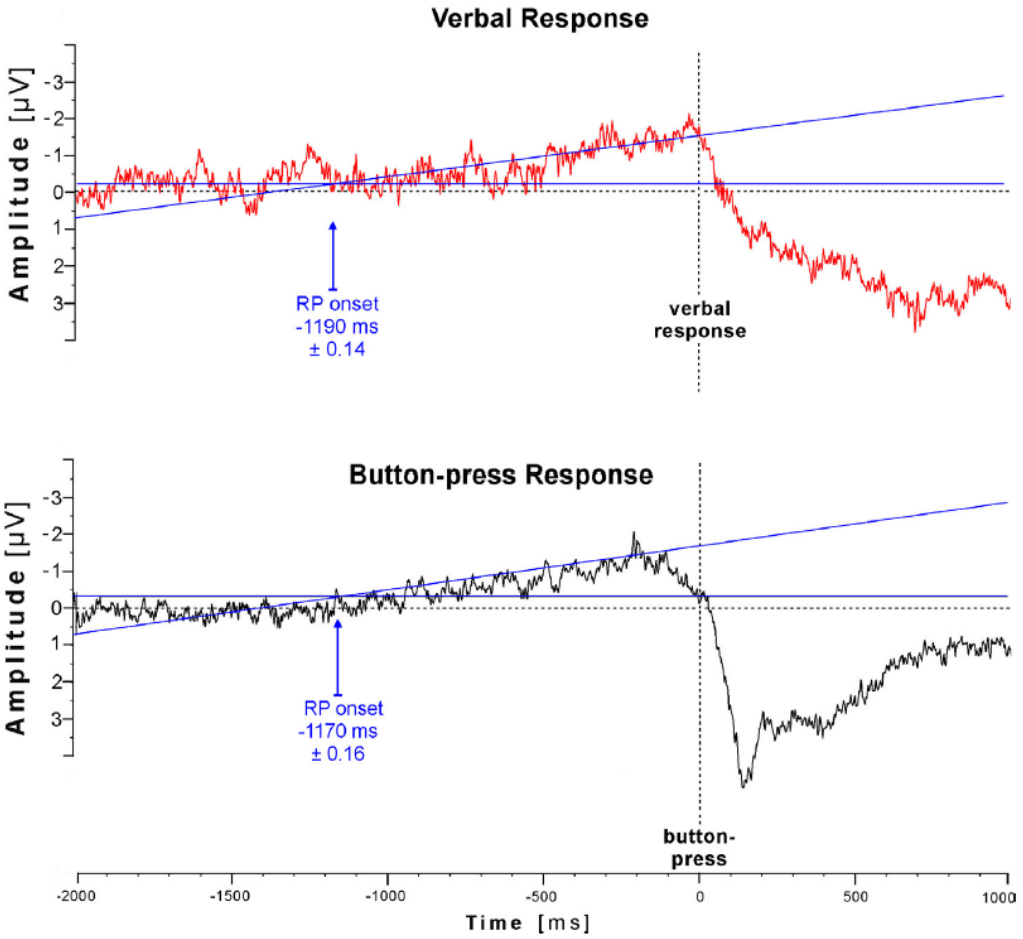
=> Reflects early anticipation of turn end

Magyari, Bastiaansen, De Ruiter, & Levinson (2014)

Turn-end anticipation (2)

Readiness Potential (RP)

- button-press
- verbal response

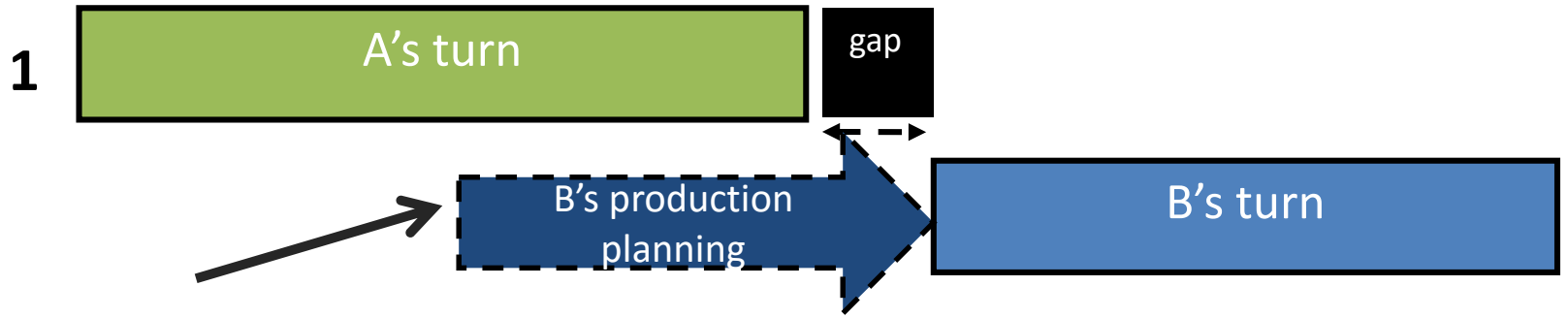


=> Reflects early anticipation of turn end

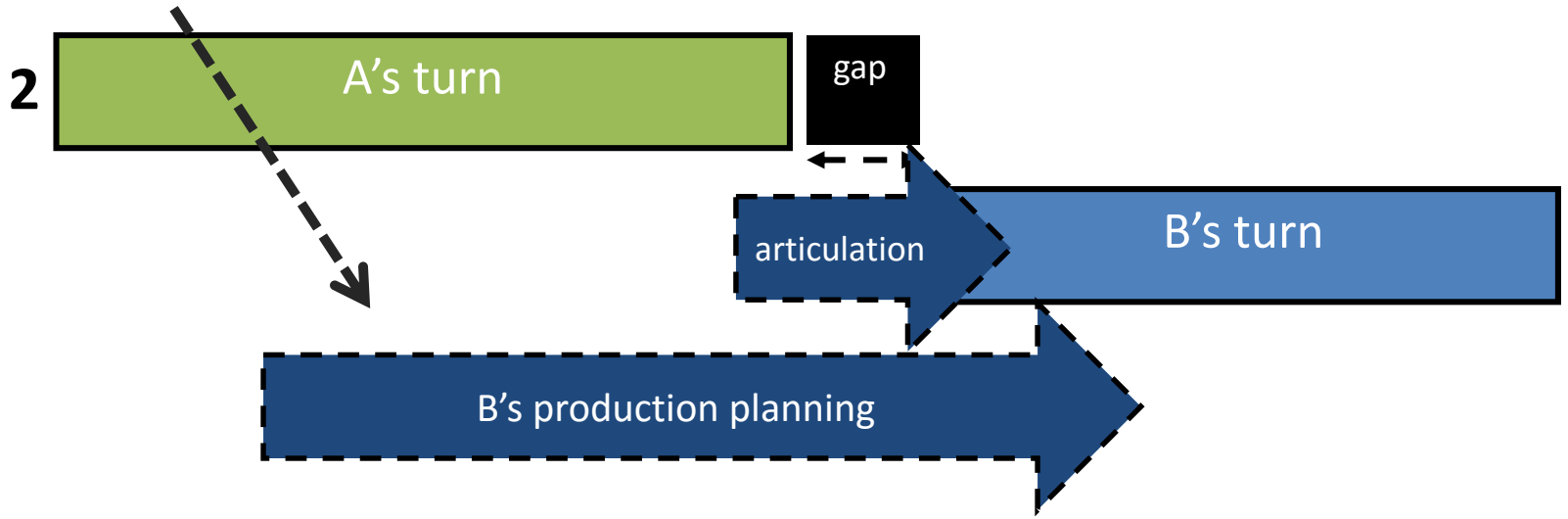
Janssen, Wesselmeier, De Ruiter, & Mueller (2014)

Models

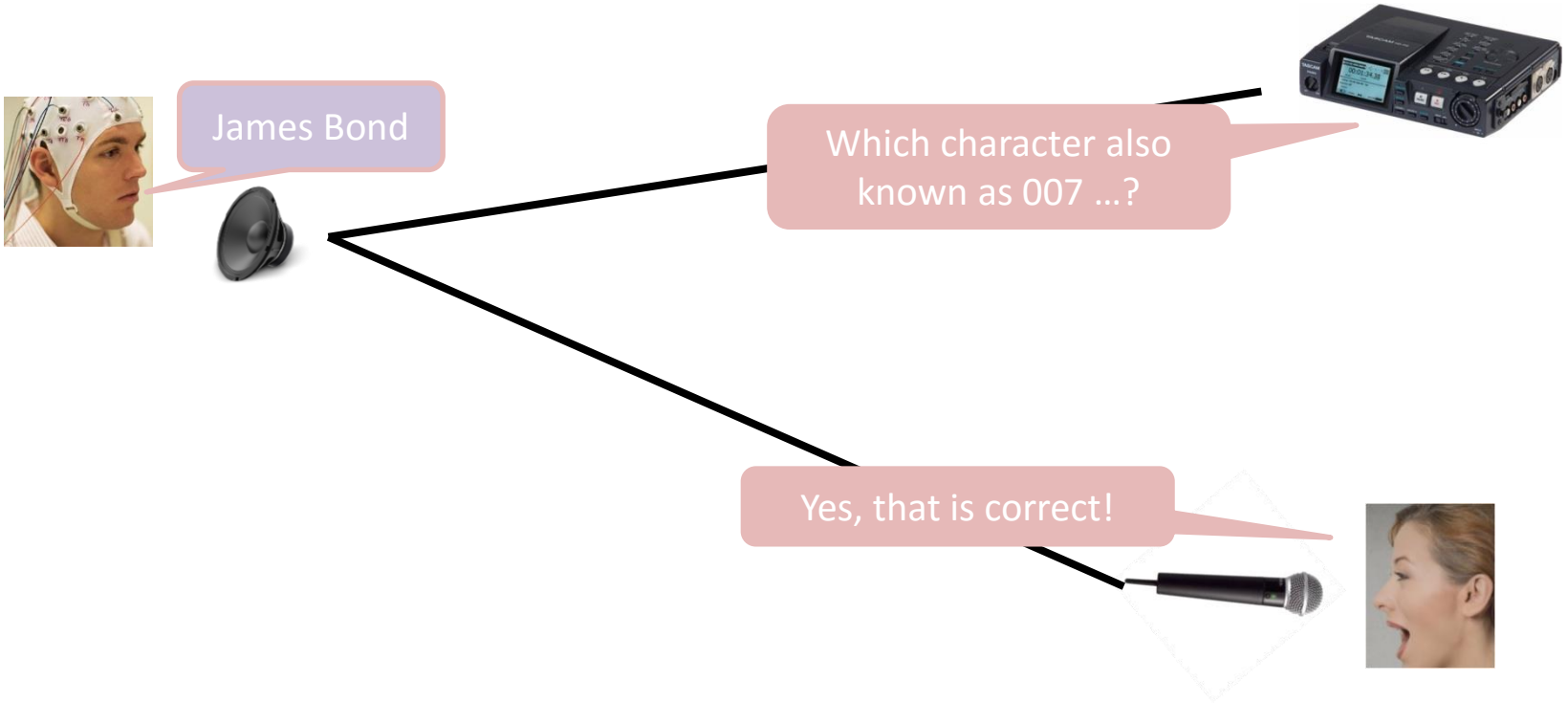
Lexical prediction



Long-range anticipation appears possible (but what is it?)



Response-planning



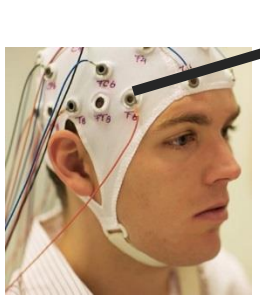
EARLY: Which character, also called **007**, appears in the famous movies?
LATE: Which character from the famous movies, is also called **007**?

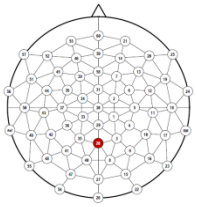
Bögels, Magyari, & Levinson (2015)

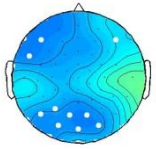
Response-planning

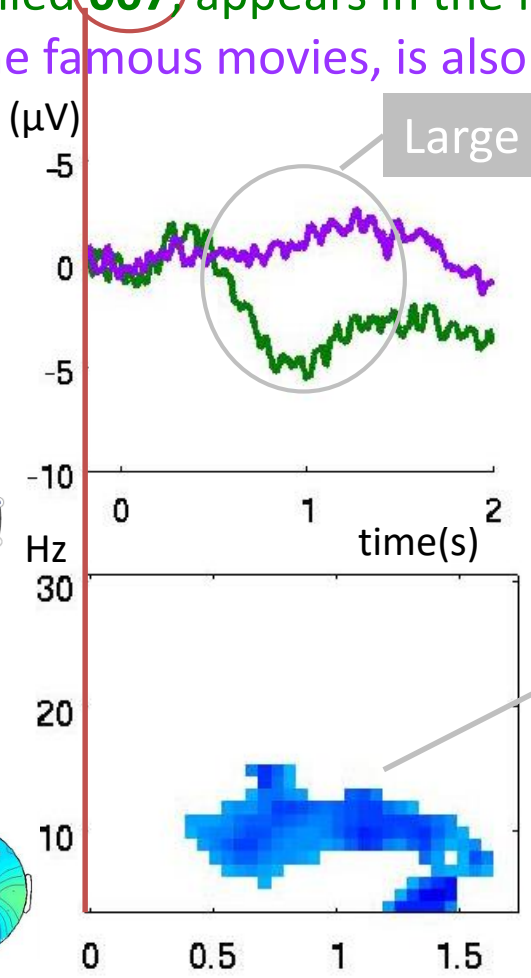
Which character, also called **007**, appears in the famous movies?

Which character from the famous movies, is also called **007**?



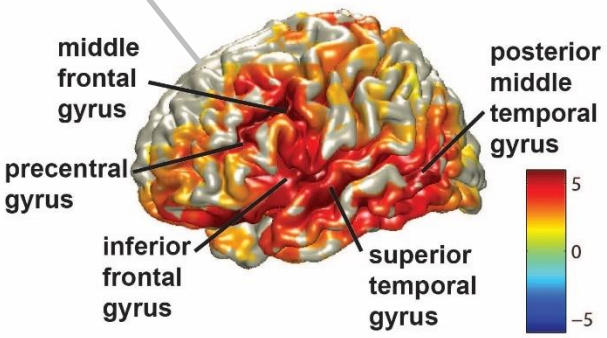






Large positivity

localized in production areas

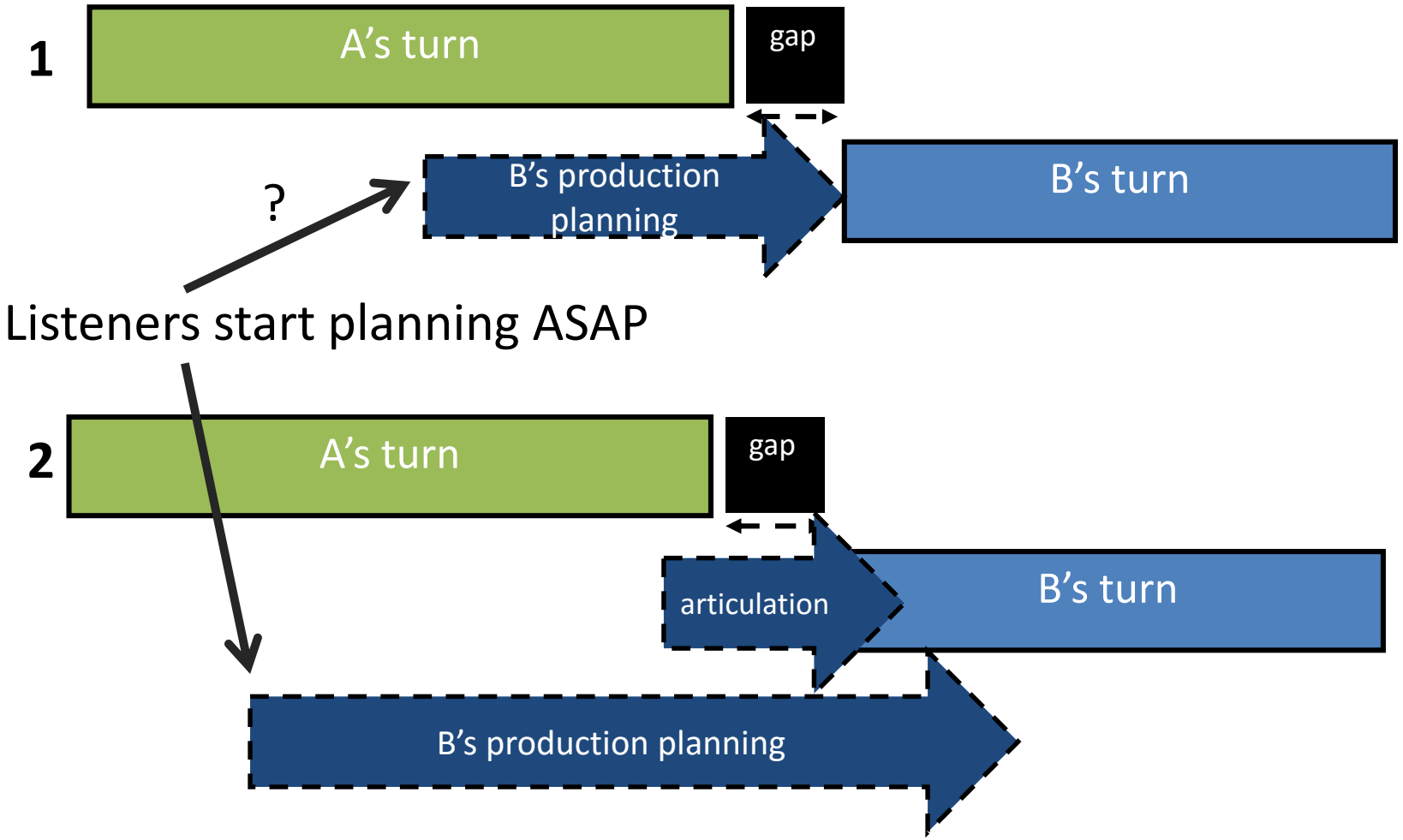


Alpha suppression

Absent/reduced in control study

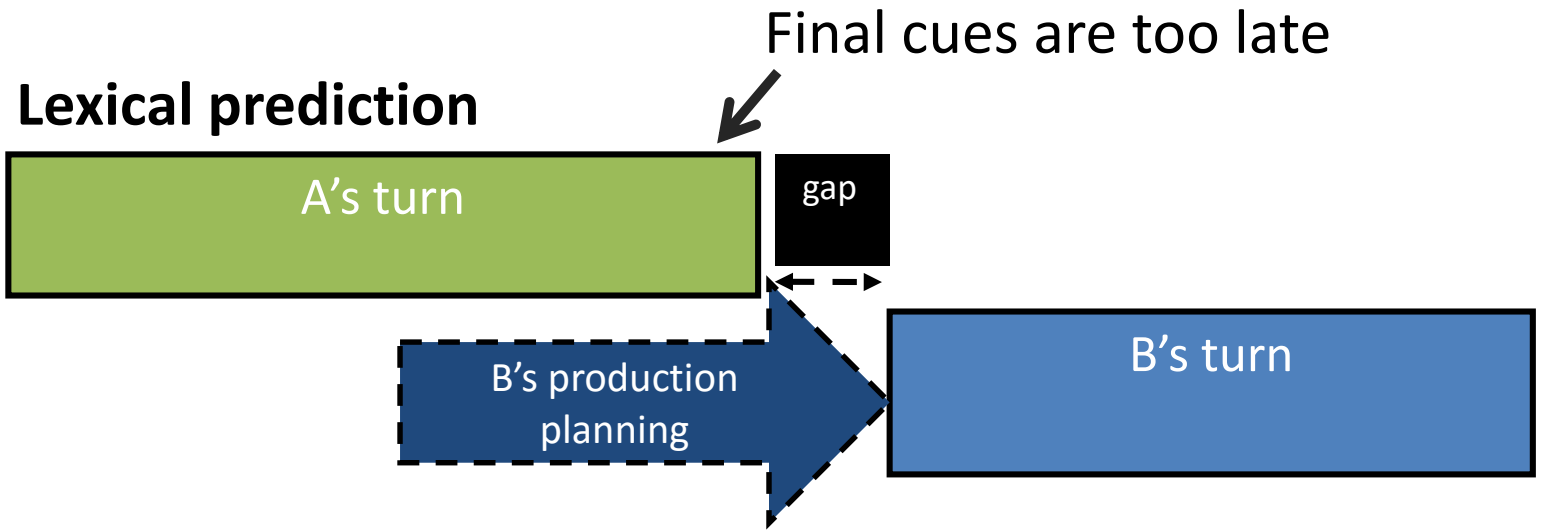
=> Listeners start planning ASAP

Models


1
A's turn
gap
B's production planning
B's turn
?
Listeners start planning ASAP
2
A's turn
gap
articulation
B's turn
B's production planning

Turn-taking &
Final prosodic cues

Are final cues too late?



How fast is reaction to turn-final cues?

De Ruiter, Mitterer & Enfield (2006)

Methods

Say ‘ja’ in reaction to stimulus offset ASAP

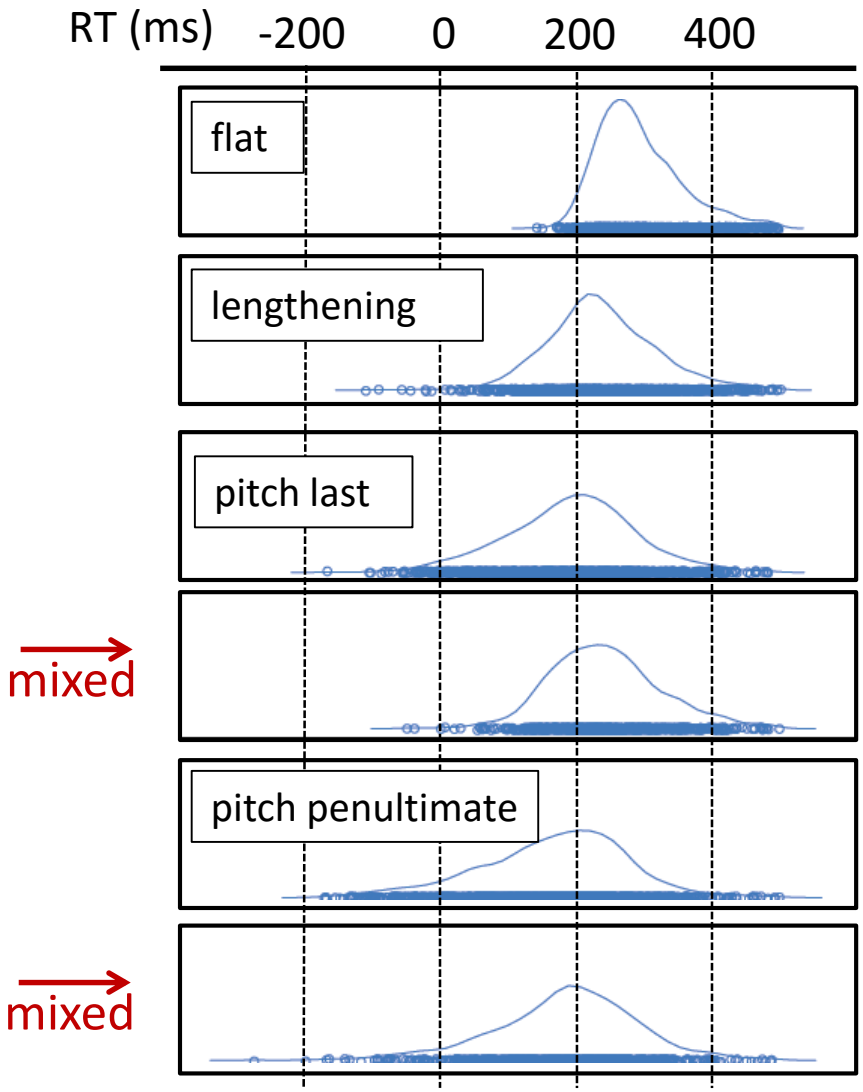
Materials (blocked presentation):

Speech-like repetitive stimulus (mamama...)

- ‘flat’: no prosodic markers
- lengthening on last syllable
- pitch change on last syllable
- pitch change on penultimate syllable
- mixed pitch block: less predictable

Torreira & Bögels (in preparation)

Results

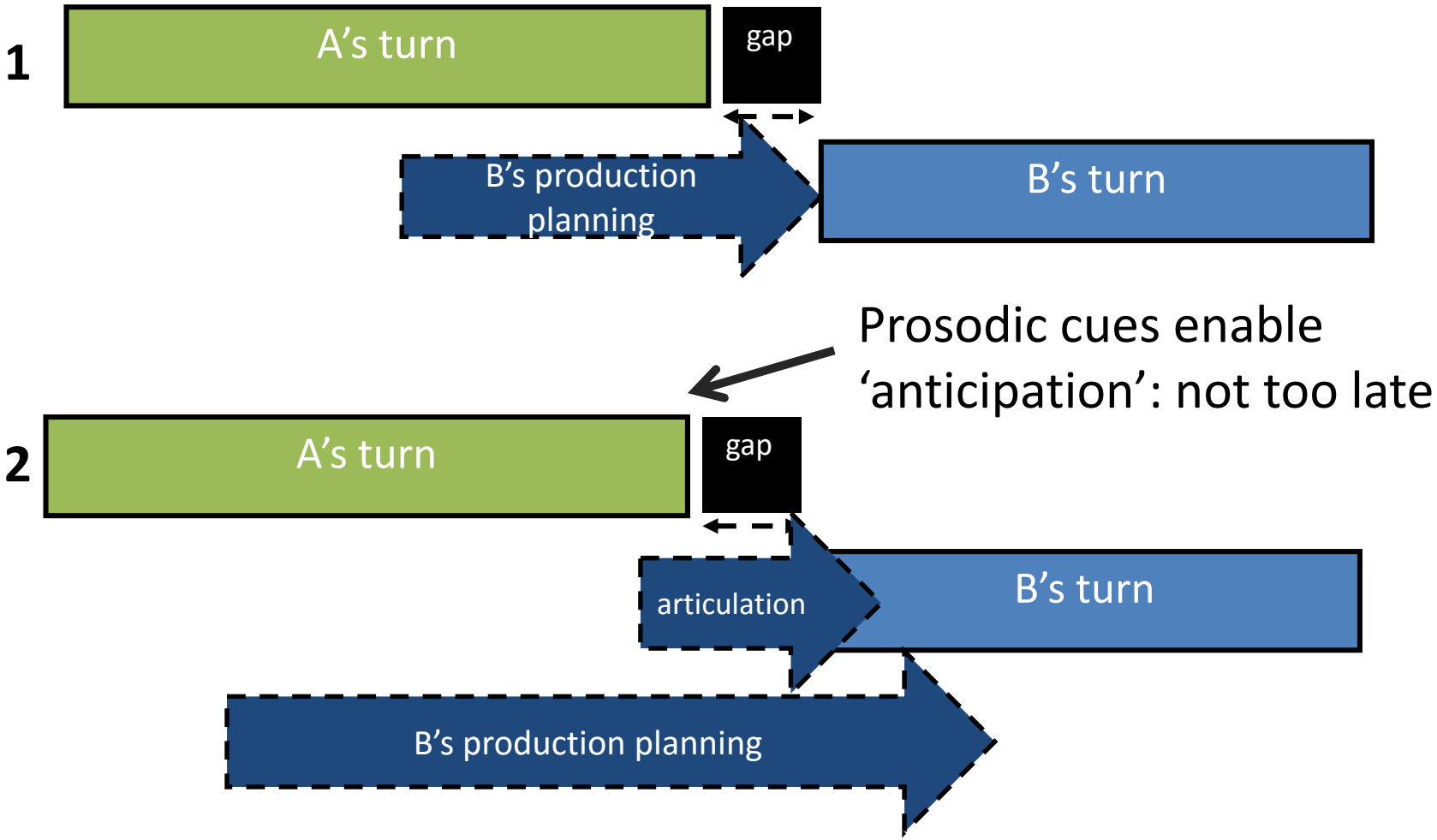


- flat: mode later than in conversation
- prosodic cues lead to faster RTs
- participants anticipate (even though asked to *react*)
- mixed block: still anticipation (less)

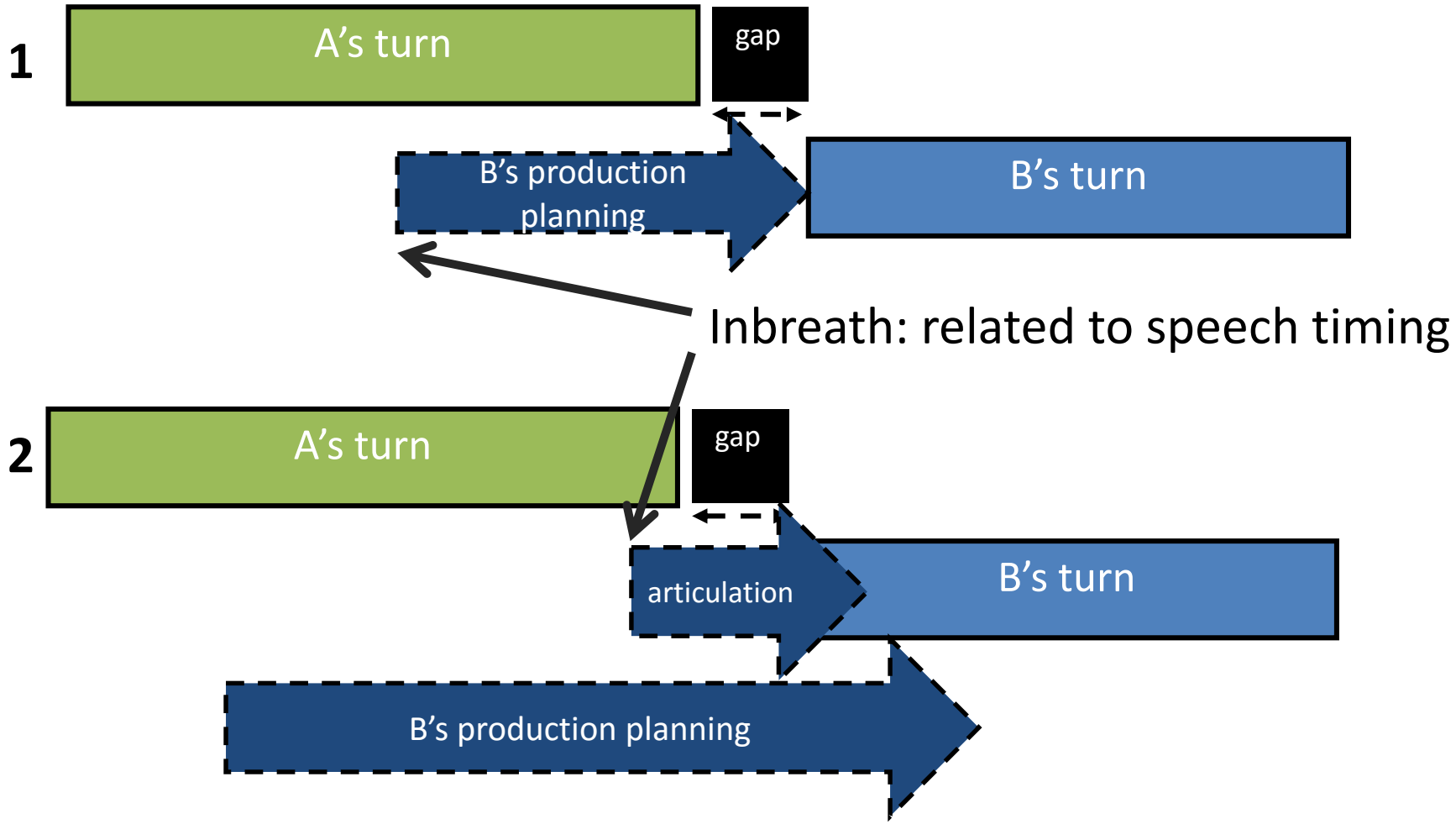
Are final cues too late?

- Prosodic cues do not seem to occur too late: they seem to enable ‘anticipation’ of the end
- ‘Anticipation’ seems to be a default strategy if possible (even if asked to react)
- Distribution of turn transitions in conversation similar to prosodic conditions
  - If content of turn is already planned; final prosodic cues could account for distribution in conversation
- A part of turn transitions in conversation could even be a reaction to silence

Final cues not too late

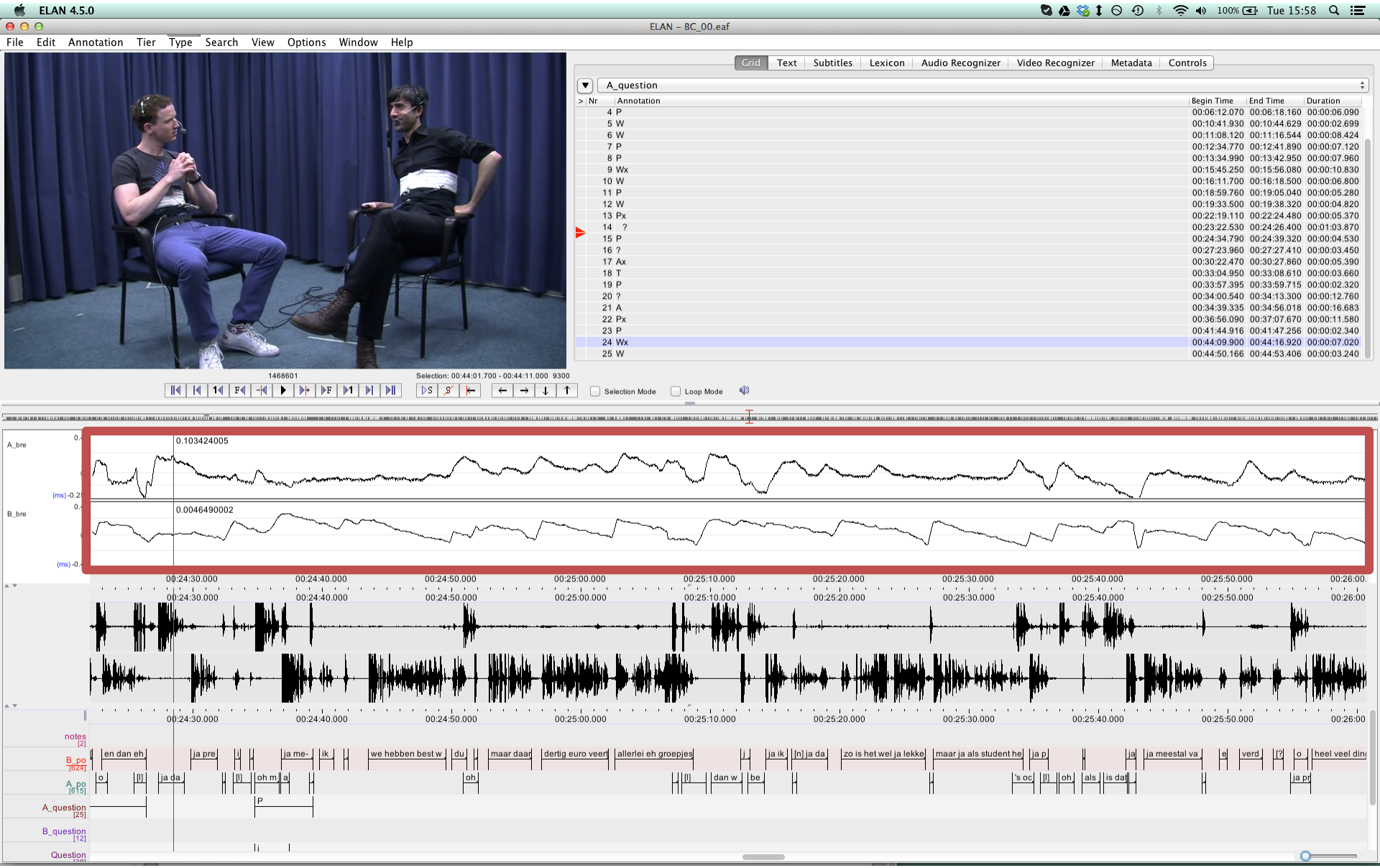


Orientation to turn-ends? Breathing



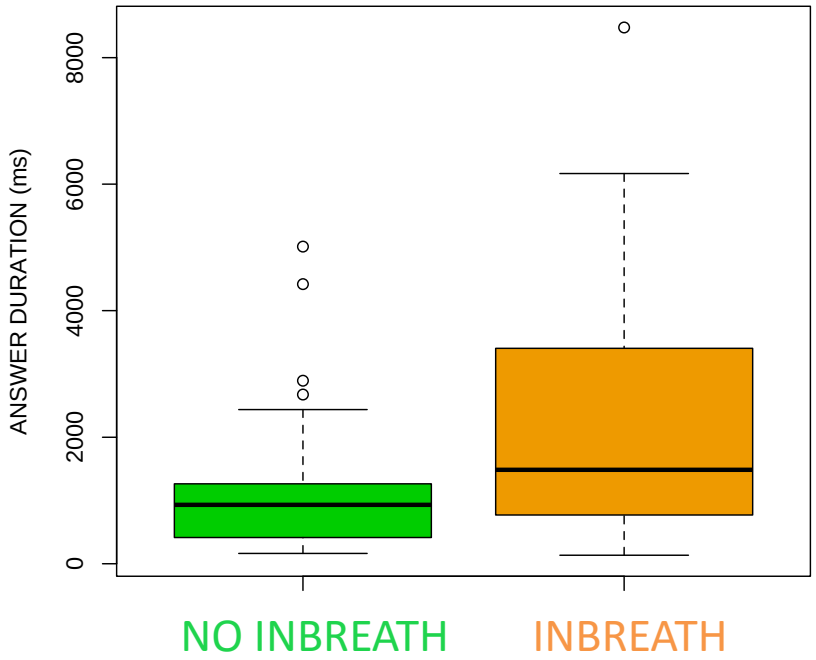
Conversational corpus with Respitrace inductive plethysmography

Torreira, Bögels, & Levinson (2015)


ELAN 4.5.0
Tue 15:58
ELAN – BC_00.eaf
File Edit Annotation Tier Type Search View Options Window Help
Grid Text Subtitles Lexicon Audio Recognizer Video Recognizer Metadata Controls
A_question
> Nr Annotation Begin Time End Time Duration
4 P 00:06:12.070 00:06:18.160 00:00:06.090
5 W 00:10:41.930 00:10:44.629 00:00:02.699
6 W 00:11:08.120 00:11:16.544 00:00:08.424
7 P 00:12:34.770 00:12:41.890 00:00:07.120
8 P 00:13:34.990 00:13:42.950 00:00:07.960
9 Wx 00:15:45.250 00:15:56.080 00:00:10.830
10 W 00:16:11.700 00:16:18.500 00:00:06.800
11 P 00:18:59.760 00:19:05.040 00:00:05.280
12 W 00:19:33.500 00:19:38.320 00:00:04.820
13 Px 00:22:19.110 00:22:24.480 00:00:05.370
14 ? 00:23:22.530 00:24:26.400 00:01:03.870
15 P 00:24:34.790 00:24:39.320 00:00:04.530
16 ? 00:27:23.960 00:27:27.410 00:00:03.450
17 Ax 00:30:22.470 00:30:27.860 00:00:05.390
18 T 00:33:04.950 00:33:08.610 00:00:03.660
19 P 00:33:57.395 00:33:59.715 00:00:02.320
20 ? 00:34:00.540 00:34:13.300 00:00:12.760
21 A 00:34:39.335 00:34:56.018 00:00:16.683
22 Px 00:36:56.090 00:37:07.670 00:00:11.580
23 P 00:41:44.916 00:41:47.256 00:00:02.340
24 Wx 00:44:09.900 00:44:16.920 00:00:07.020
25 W 00:44:50.166 00:44:53.406 00:00:03.240
1468601
Selection: 00:44:01.700 - 00:44:11.000 9300
Selection Mode
Loop Mode
A_bre
B_bre
0.103424005
0.0046490002
00:24:30.000
00:24:40.000
00:24:50.000
00:25:00.000
00:25:10.000
00:25:20.000
00:25:30.000
00:25:40.000
00:25:50.000
00:26:00.
notes [2]
B_po [624]
A_po [615]
A_question [25]
B_question [12]
Question
en dan eh
ja pre
ja me-
ik
we hebben best w
du
maar daar
dertig euro veert
allerlei eh groepjes
ja ik
[n] ja da
zo is het wel ja lekke
maar ja als student he
ja p
ja
ja meestal va
verd
heel veel din
o
[l]
ja da
oh m
a
oh
[l]
dan w
be
's oc
oh
als
is dat
ja pr
P

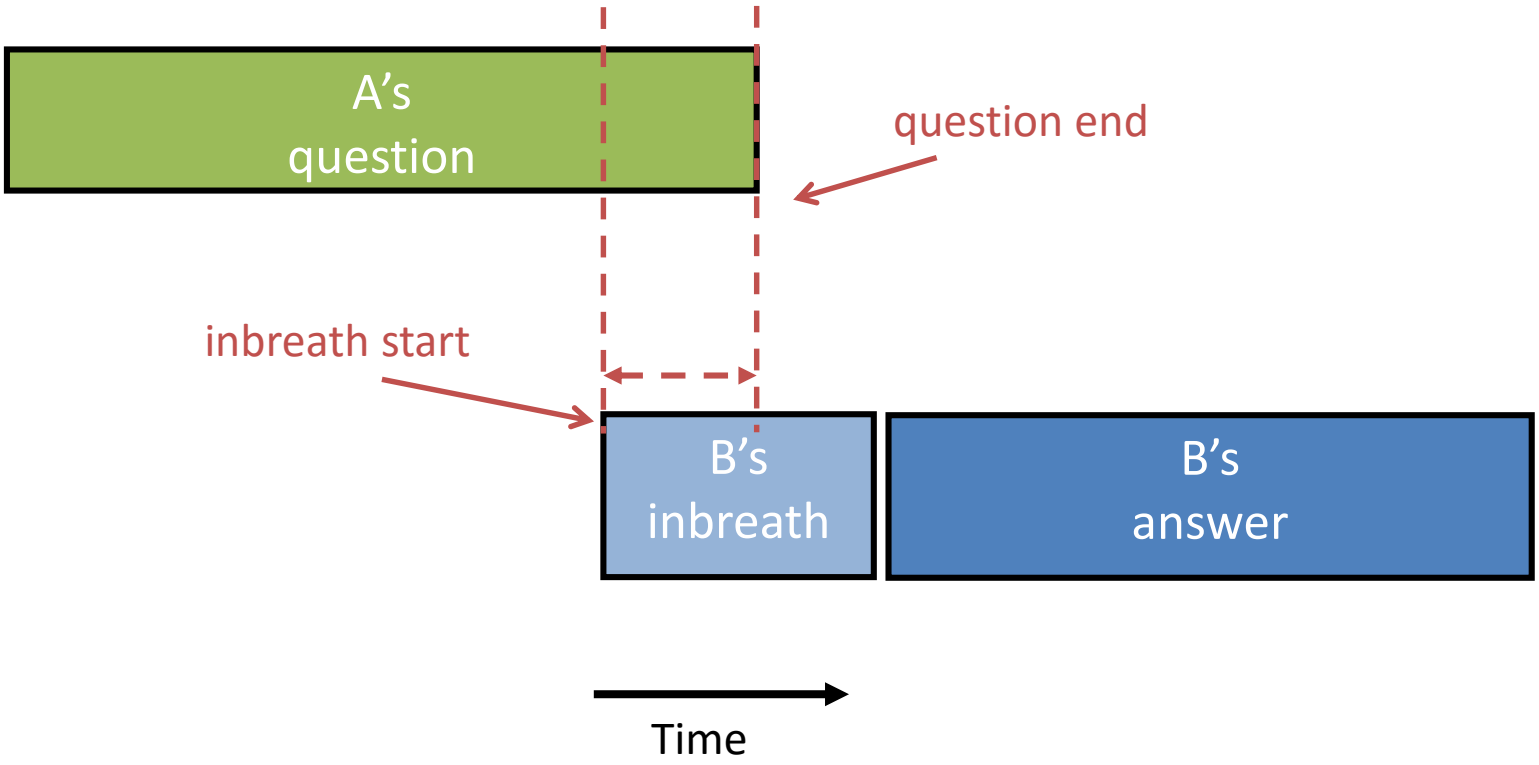
Answer duration & inbreaths

Question-answer pairs

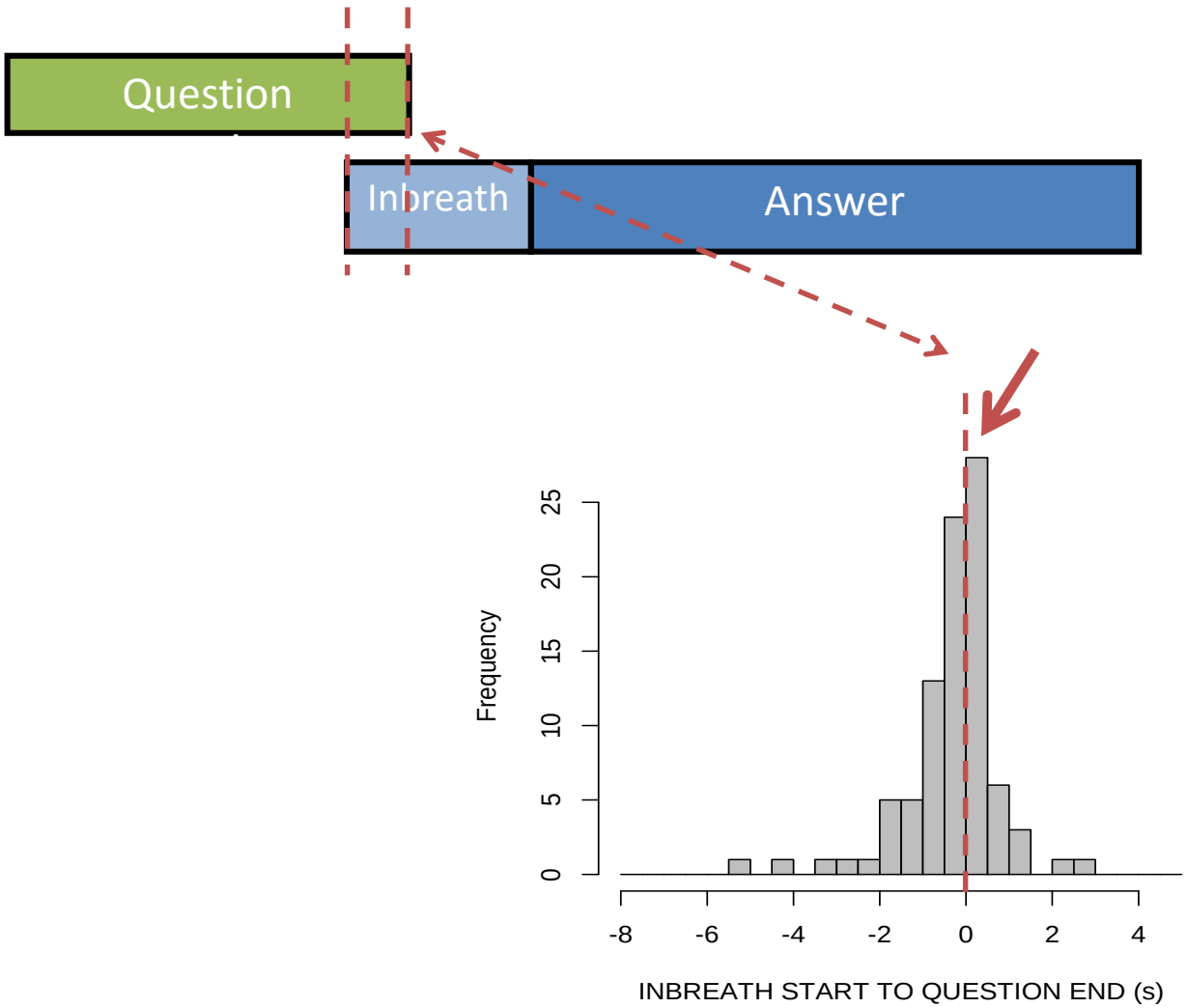


=> Inbreaths related to planning

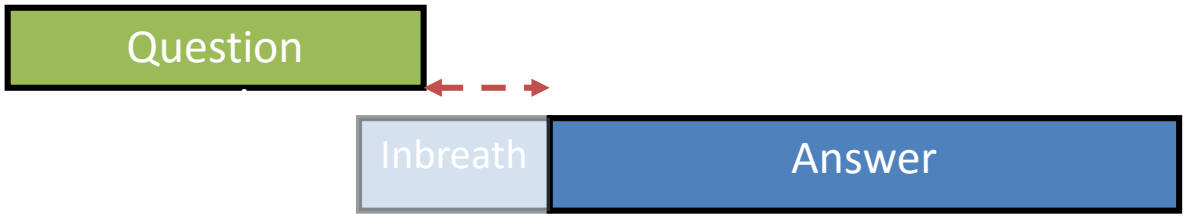
Timing relative to question end

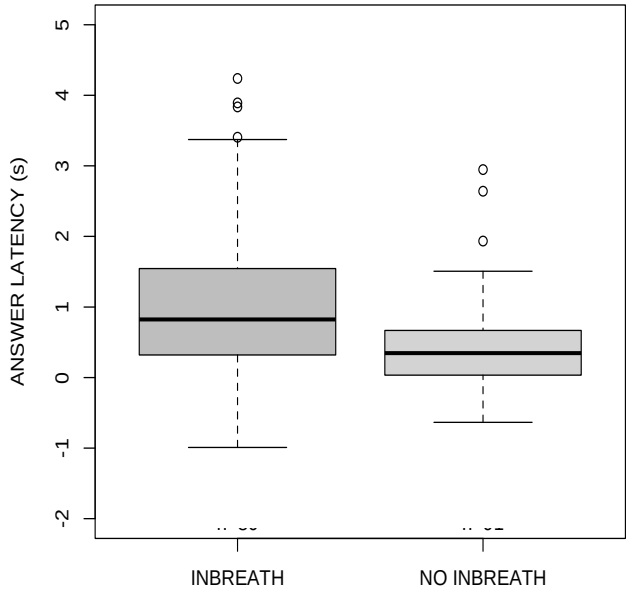


Inbreath timing



Answer timing





Conclusions

Inbreaths are more likely to occur before longer answers

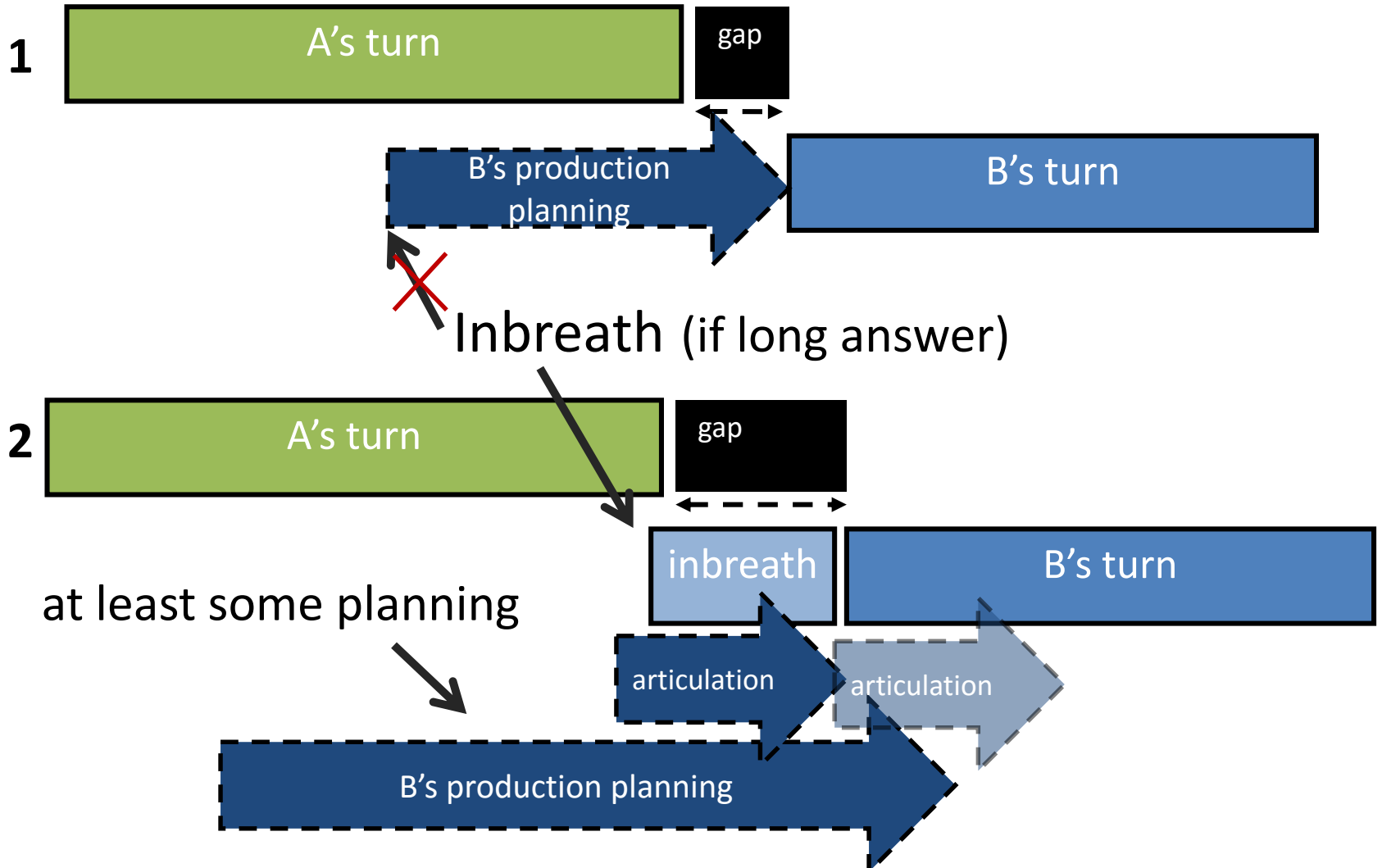
> breathing behavior can be informative about speech planning in conversation

Speech inbreaths before answers appear to be timed to the end of questions

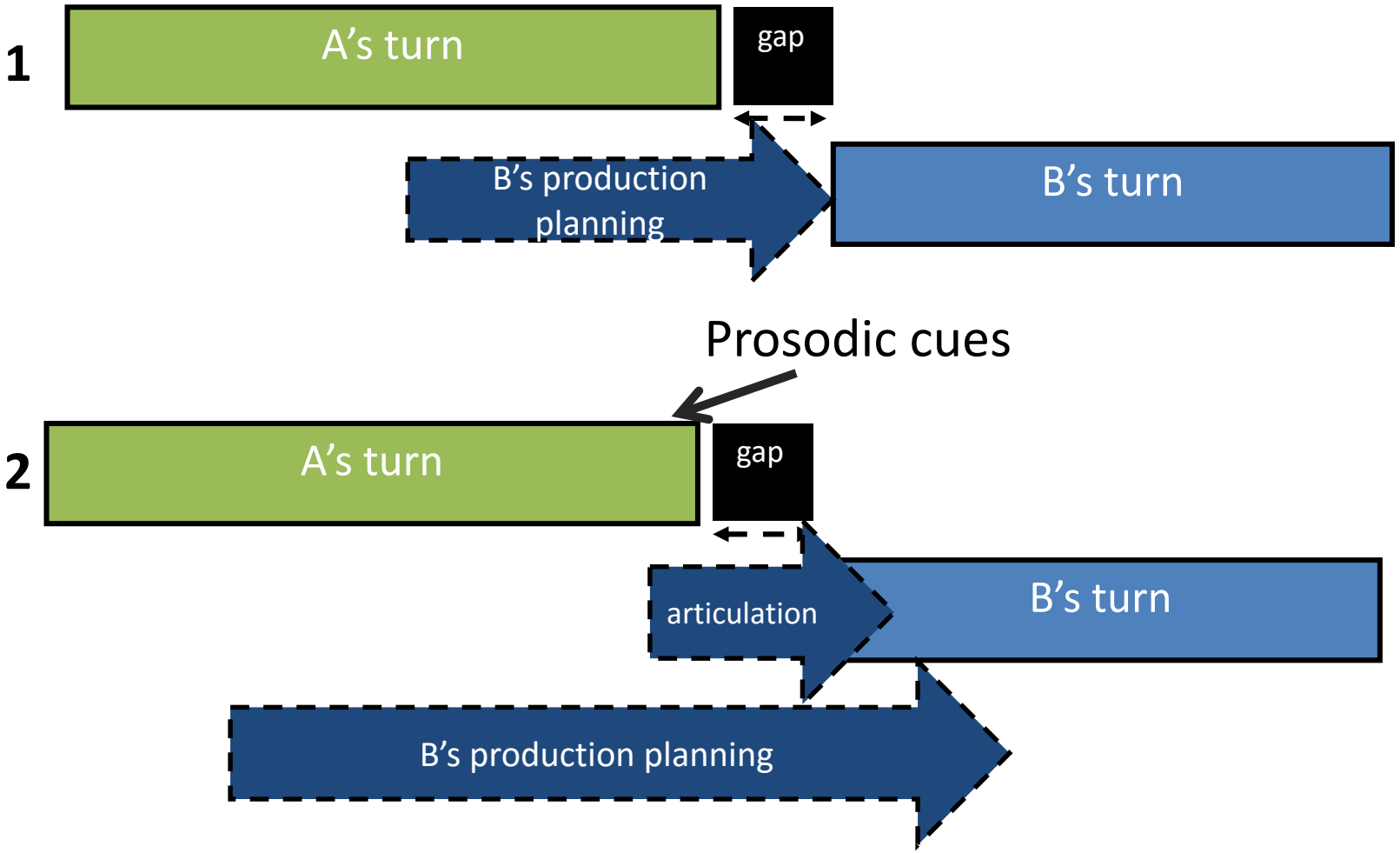
> consistent with the idea that *articulation* is timed to turn ends (*planning* might start earlier)

> answers with an inbreath occur later

Listeners orient to turn-ends



Use of final prosodic cues?



Prosody & turn-boundaries

- Turn-boundaries coincide with:
  - Intonational completion (nuclear pitch accent, boundary tone, final lengthening)
  - Lexico-syntactic completion

- Observational studies: correlation
- Offline experiments: same as online?

Caspers (2001), Ford & Thompson (1996), Gravano & Hirschberg (2011), Hjalmarsson (2011), Local & Walker (2013), Wells & McFarlane (1998)

Online experiment

Extraction of turns from a corpus of conversational Dutch

Button-press task:

- Participants anticipate the end of turns
- Conditions:
  - Original (no manipulation)
  - No pitch (flattened)
  - No words (low-pass filtering)

De Ruiter, Mitterer & Enfield (2006)

Online experiment

Results:

No pitch = Original

No words < Original

Conclusions:

- Lexico-syntactic information is necessary and possibly sufficient for turn-end projection
- Intonation is neither necessary nor sufficient

  But pitch ≠ intonation!
  Participants may have used other cues to intonational phrasing in the experiment (e.g. final lengthening)

De Ruiter, Mitterer & Enfield (2006)

Prosodic cues to turn-taking

RA interviewed participants via microphones and headphones

Short and long questions from a script embedded in otherwise free interview:

e.g., *So you are a student?*

So you are a student here at the Radboud University?

Same lexico-syntax: end should be ambiguous

Bögels & Torreira (2015)

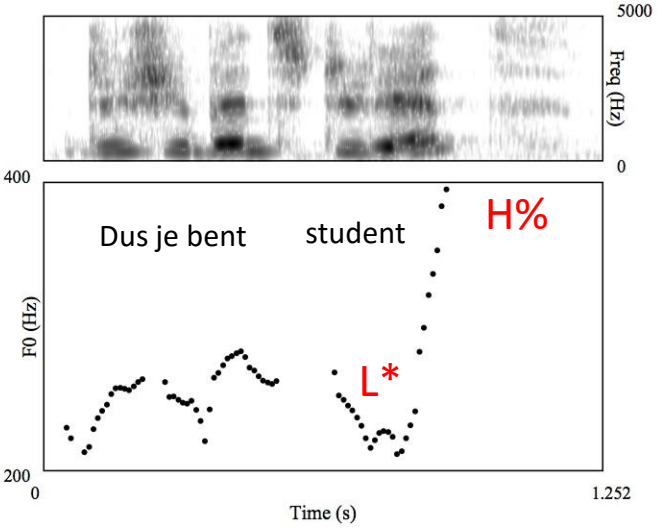
Response times

- Long questions: not many overlaps and none close to syntactic completion point

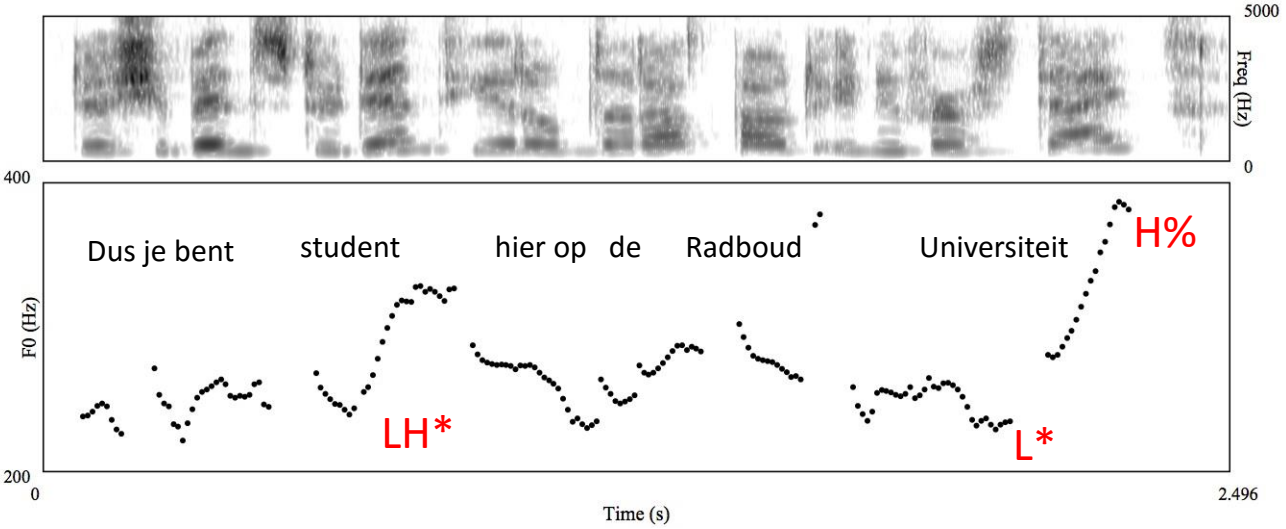
- Short questions: close to the end; 18% gap < 200 ms (no reaction)

Acoustic measurements

SHORT

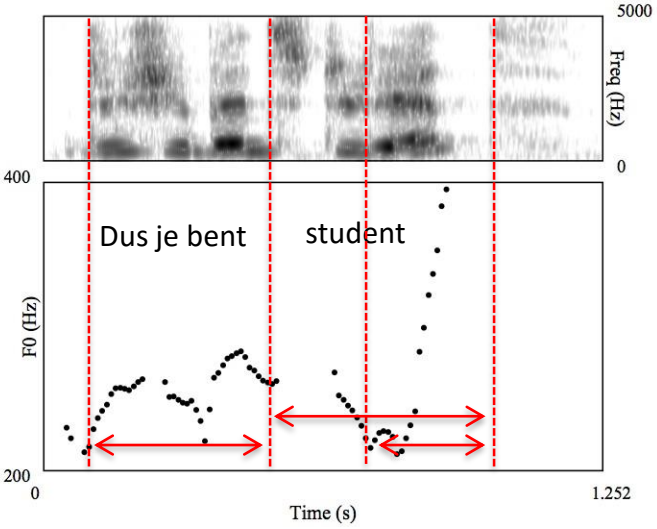


LONG



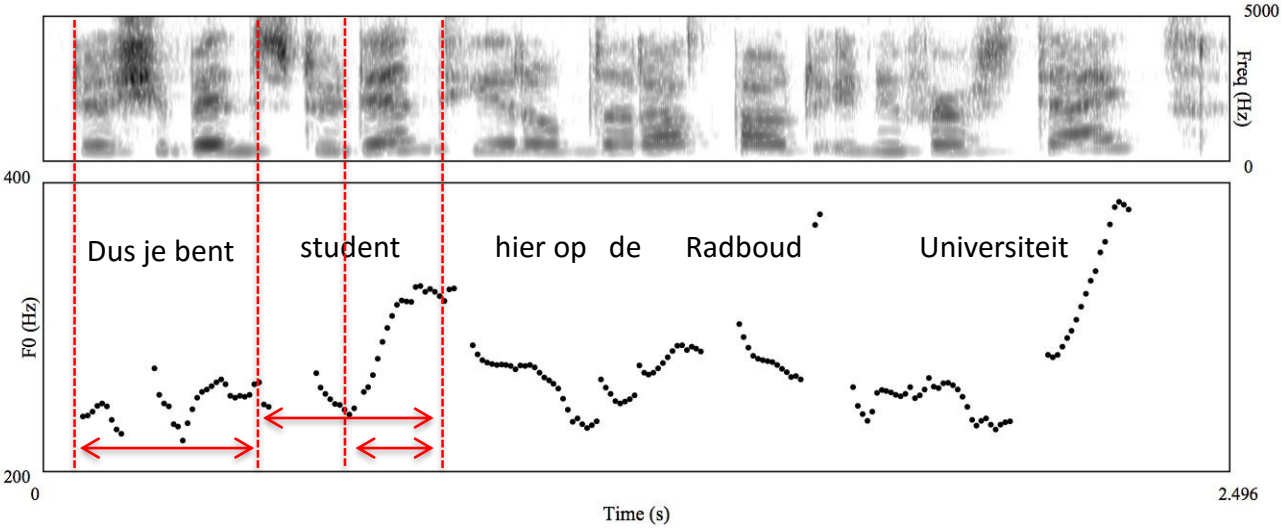
Acoustic measurements

SHORT



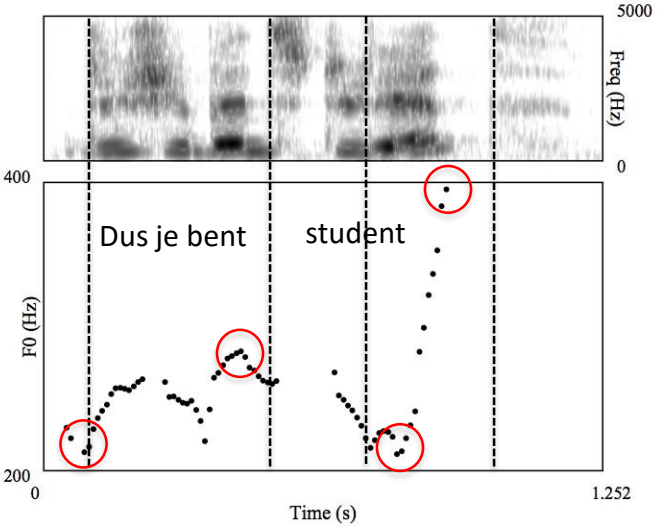
- Duration of final syllable
- Duration of final word
- Duration of preceding words

LONG



Acoustic measurements

SHORT

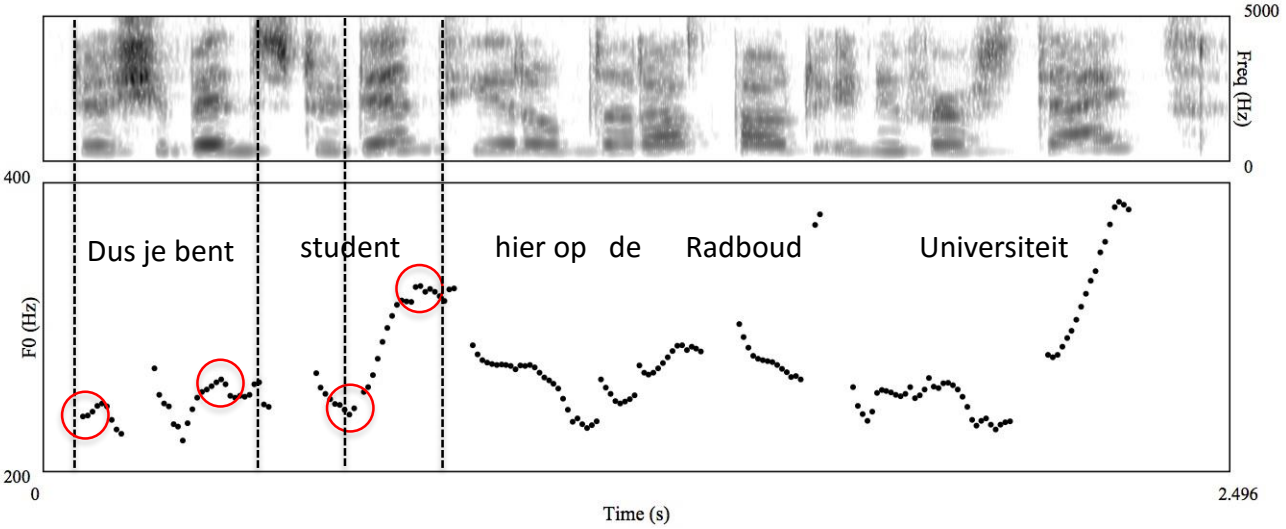


- Duration of final syllable
- Duration of final word
- Duration of preceding words

- F0 valleys

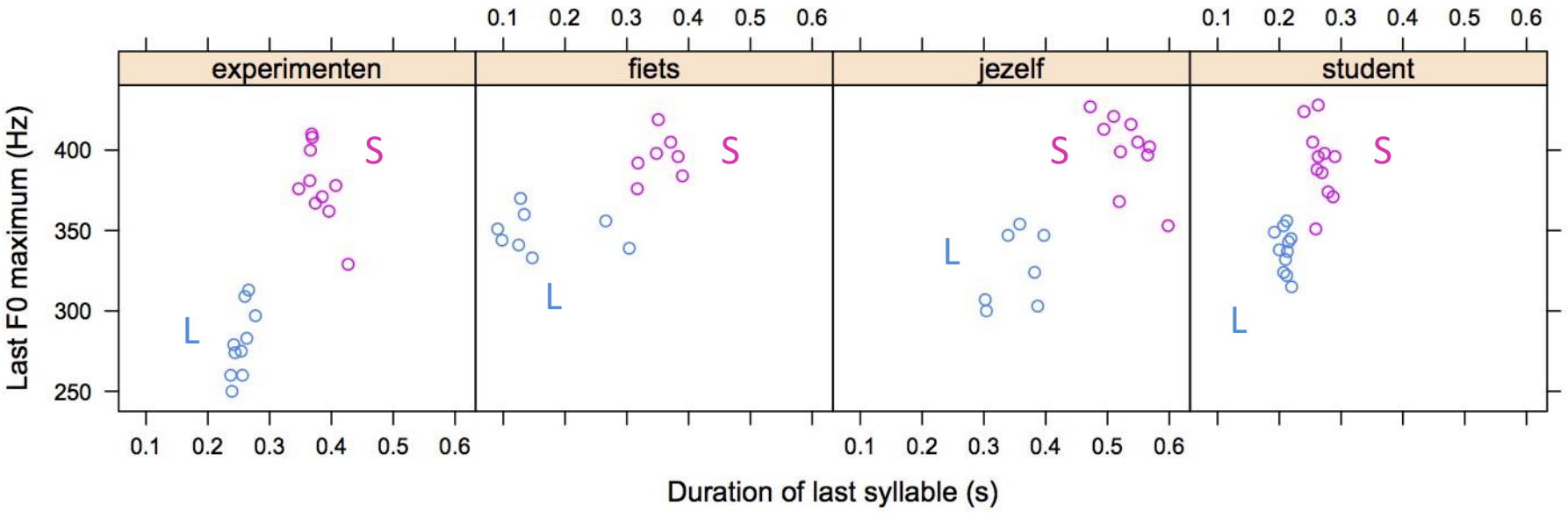
- F0 peaks

LONG

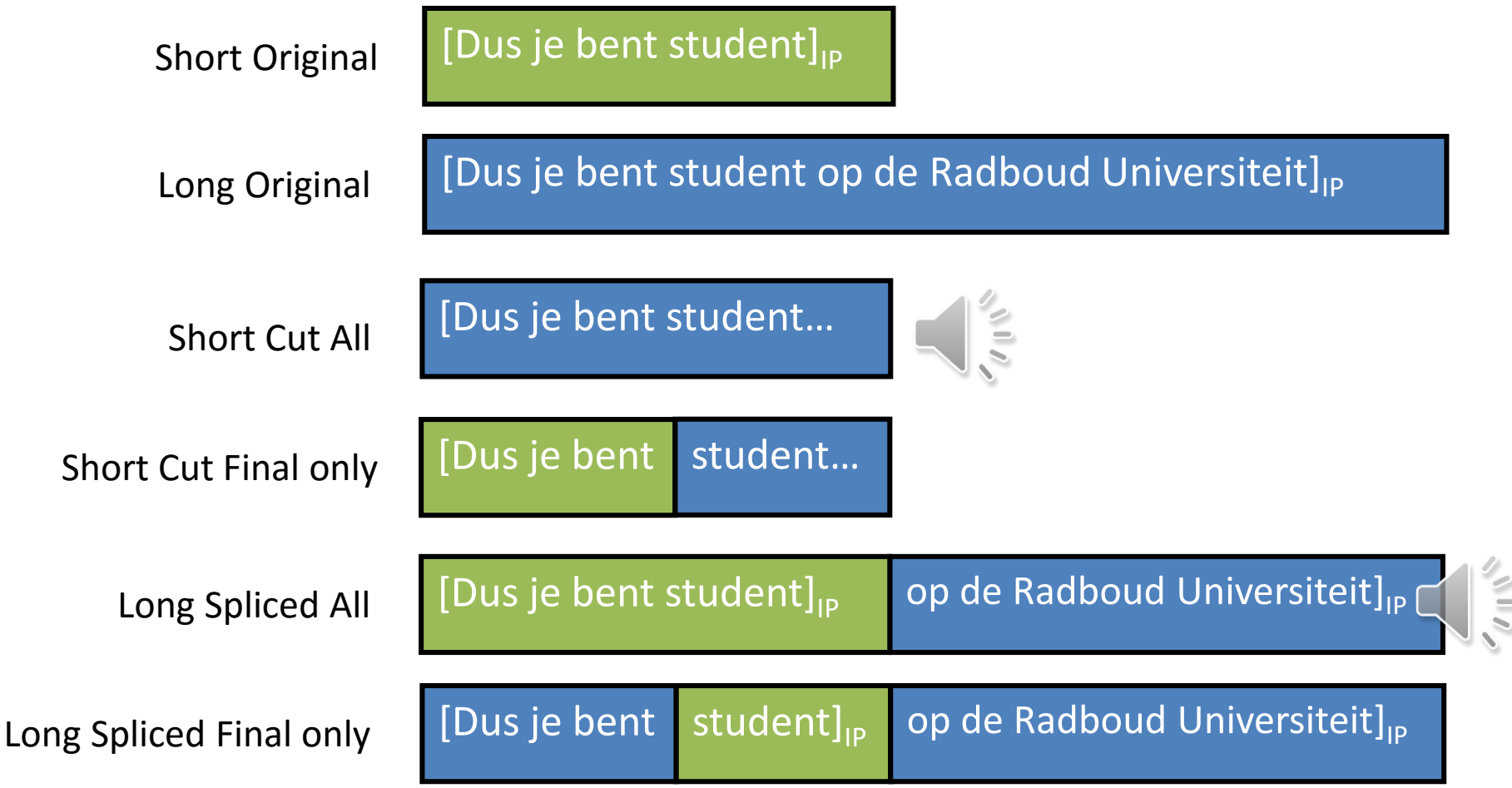


Findings

- No consistent differences before ‘final’ word
- Clear differences in ‘final’ word:

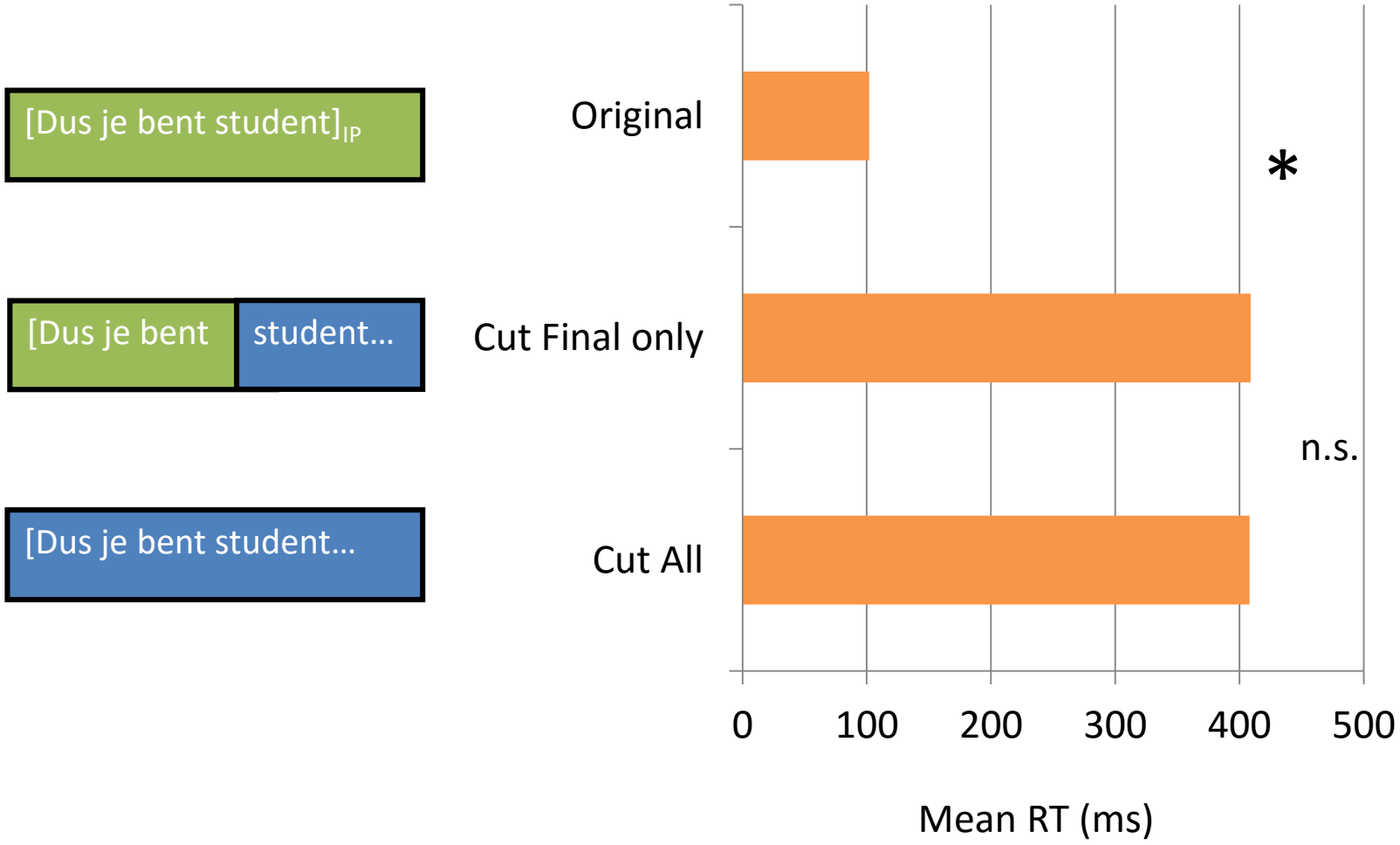


Experiment: Stimuli

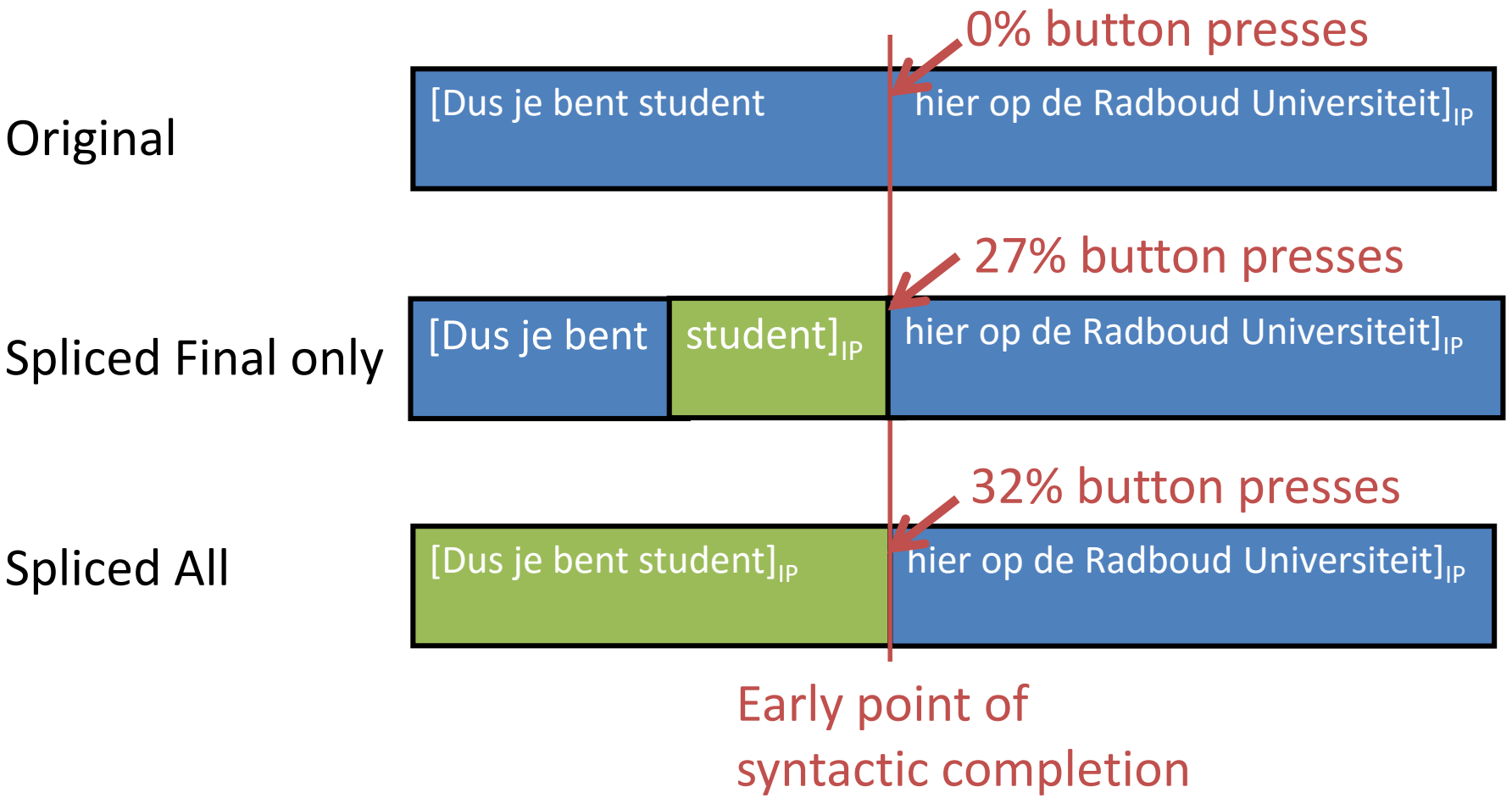


Only 2 manipulated items per participant

Results: short stimuli



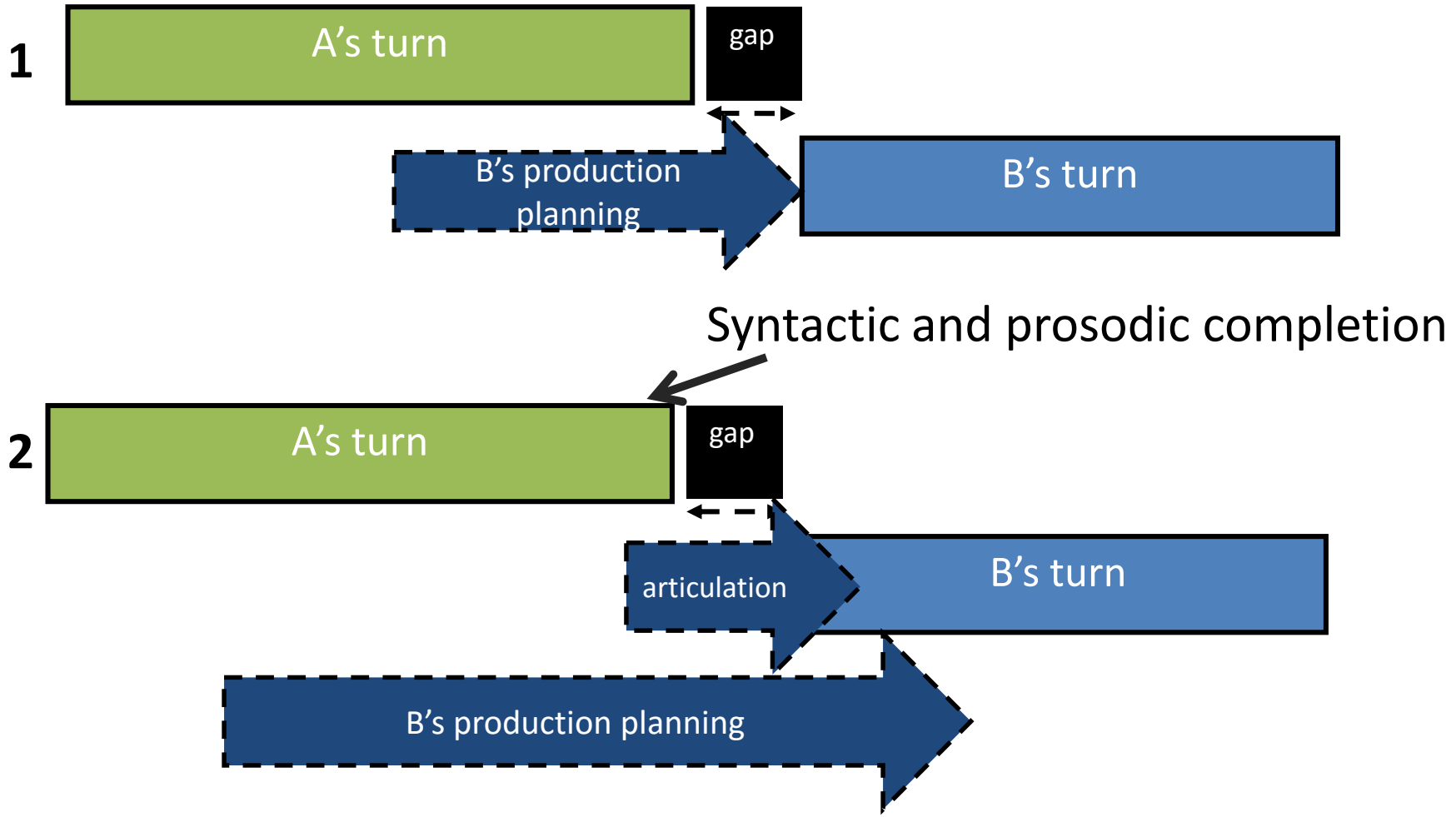
Results: long stimuli



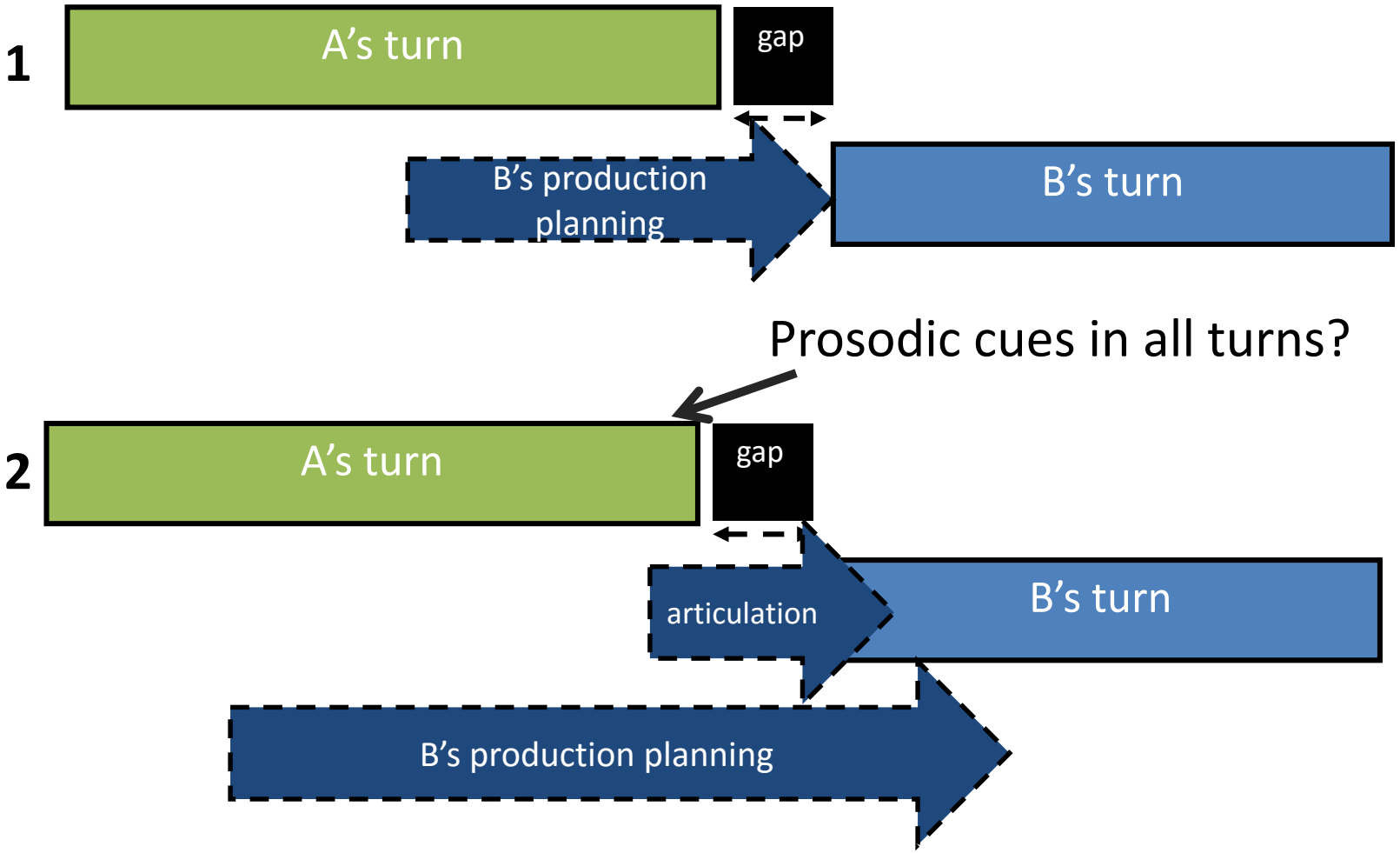
Discussion

- Listeners use prosodic cues to determine turn-ends
  - late button-presses if no IPh boundary at the end
  - some button-presses if IPh boundary in the middle
- Final cues appear most important
  - acoustic analyses
  - same effects if only last word was manipulated

Use of final prosodic cues



Generalizability



Methods

Materials

- 96 turns from spontaneous telephone calls (no overlap)
- Naïve readers identified plausible points of syntactic completion
  - See transcripts of turns word-by-word (1 s.)
  - Task: press button when see last word

=> 35 *early plausible syntactic completion points:*

10+/24 readers thought this was the last word

Prosodic analysis

Annotation of prosodic cues at these points:

- Sentence accent
- Salient final lengthening
- Salient phrase-final pitch movement

Cues are not independent

- When no accent: no melody & no lengthening (2)

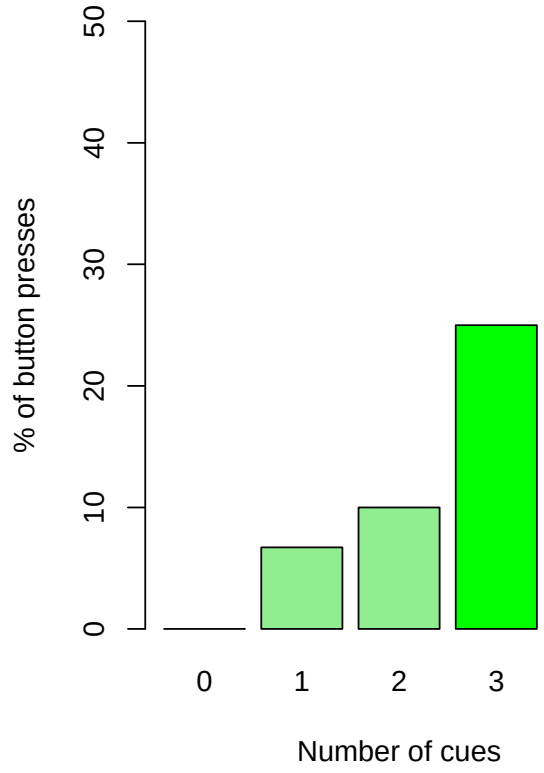
Experiment

Auditory task

- Button-press
- All non-manipulated turns

Number of cues

Window [-250, 250 ms] around plausible points of syntactic completion => how many button-presses?

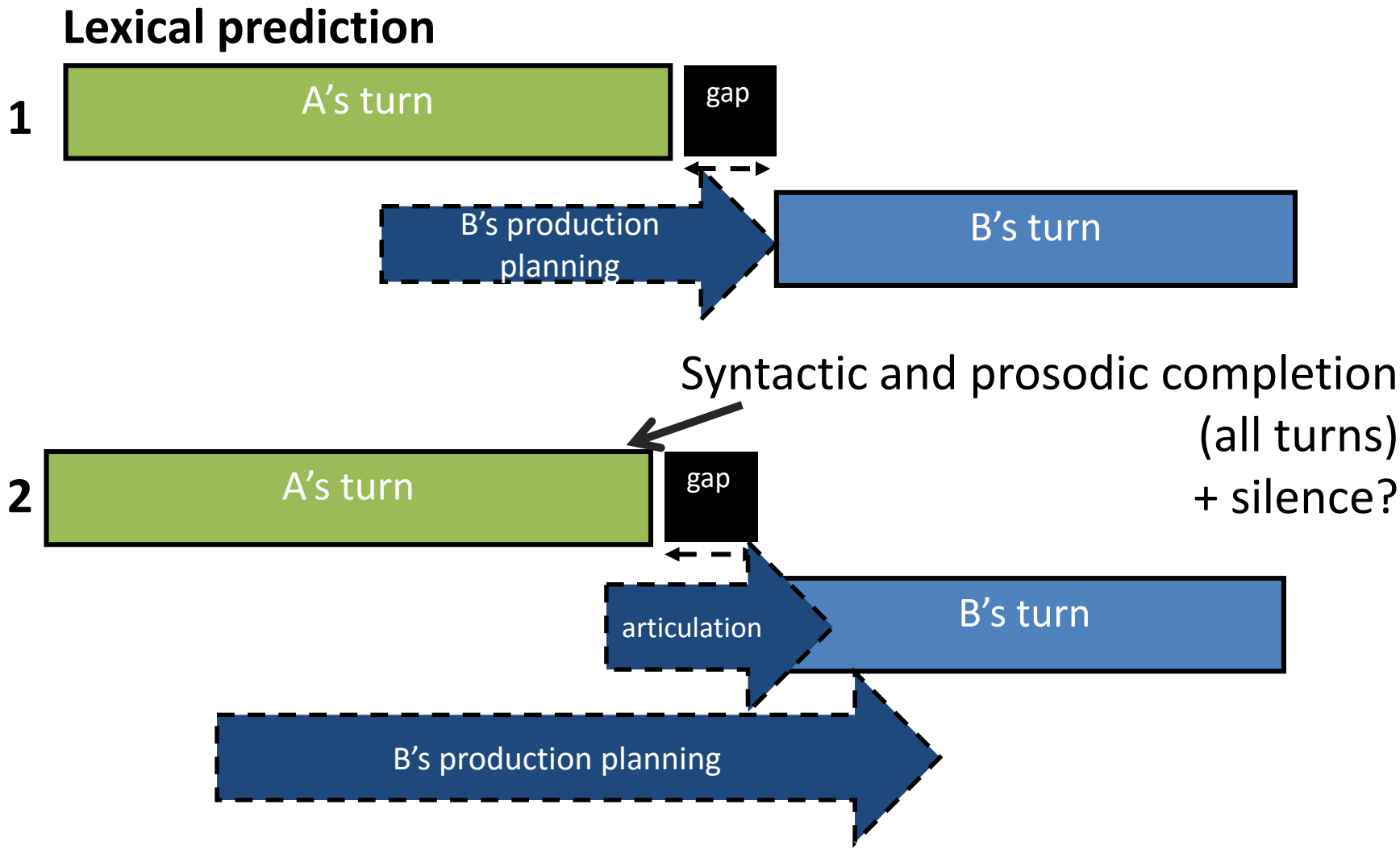


More button-presses if more prosodic cues are present ($p < .001$)

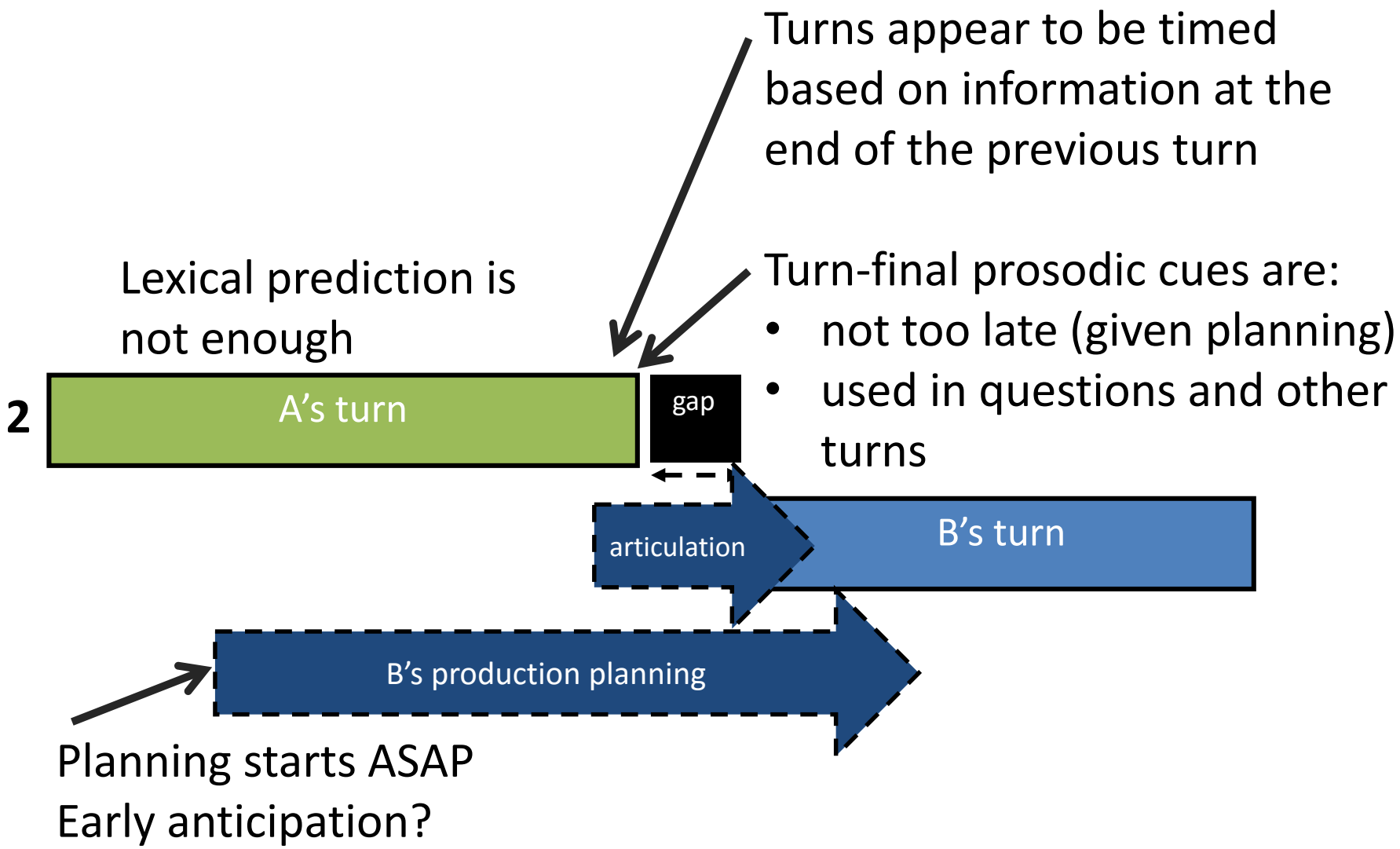
Conclusions

- Plausible early syntactic completion points are abundant (36%)
  - Not all also prosodically complete!
- Prosodic cues – not syntactic completion alone – predict anticipatory button-presses
- Still < 30% button-presses even with all cues
  - Silence used some of the time?
  - Anticipation + inhibition?

Prosodic cues are used



Conclusions turn-taking

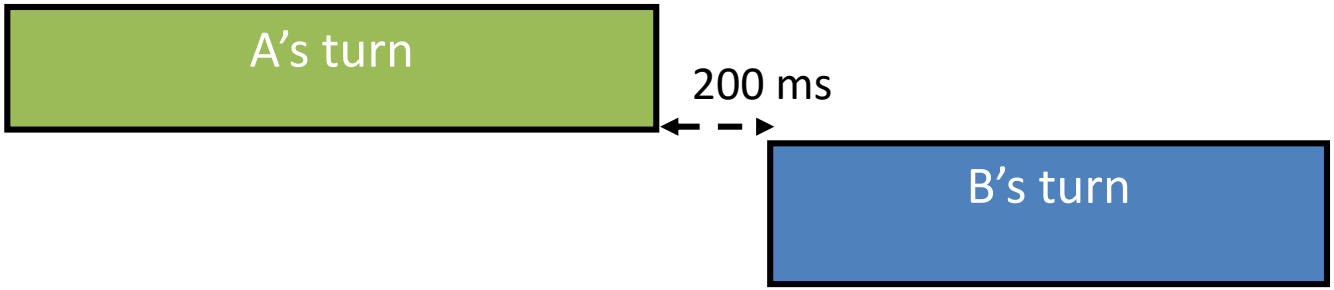


Some future directions

- Early planning in overlap with listening?
  - Preliminary evidence for trade-off between comprehension and production planning
- More natural task: button-press => answer
- Importance of different final cues: strength
- Turn-keeping cues, e.g.
  - Rush-through
  - Hesitation

Turn-taking & Silence

Long silences



Some gaps are long. Why?

e.g., *dispreferred* responses

A: *I could come to you right now?*

B: *Sure.* 'accepting' (preferred)

B': *Well, I have to work.* 'declining' (dispreferred)

(e.g., Kendrick & Torreira, 2015)

Silence & anticipation

Off-line experiments e.g., Roberts, Francis & Morgan (2006)

- Listen to request + positive response with different gap lengths
- Willingness to comply with a request judged smaller for longer silence

Online expectations affected by gap length?

⟹ EEG

Research Question

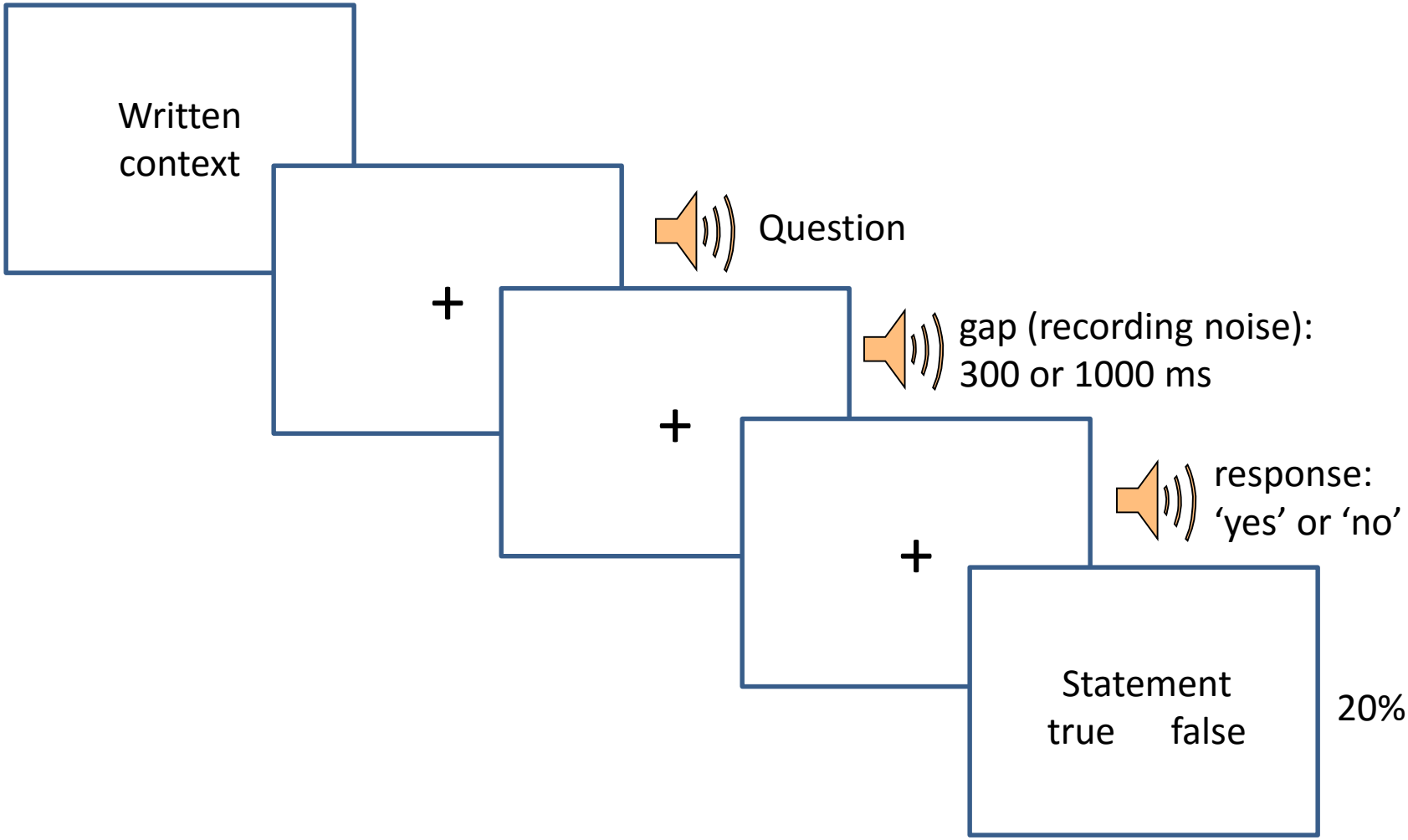
Do interlocutors anticipate the type of response (preferred or dispreferred) based on the duration of inter-turn silence alone?

Bögels, Kendrick & Levinson (2015)

Methods

- Turns from a Dutch spoken corpus (CGN): telephone conversations between friends and acquaintances
- 120 questions: requests, invitations, proposals, and offers
- 60 responses from elsewhere in the corpus
  - preferred responses: *ja* (‘yes’)
  - dispreferred response: *nee* (‘no’)
- Two response timings
  - 300 ms gap
  - 1000 ms gap

Procedure

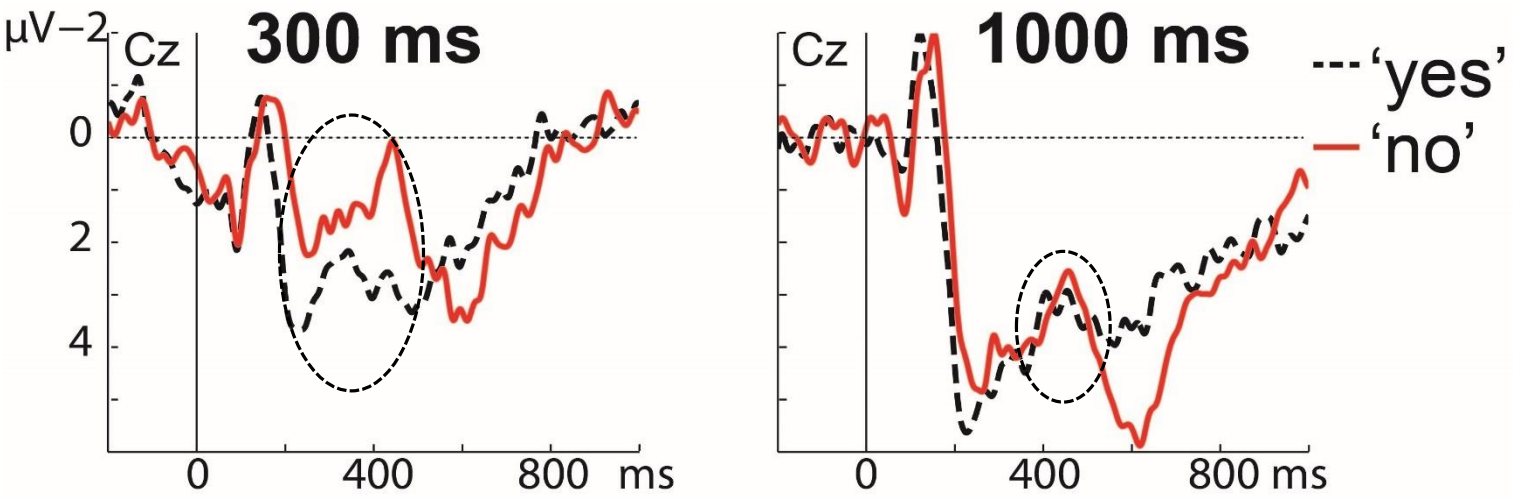

Written context
+
Question
+
gap (recording noise): 300 or 1000 ms
+
response: ‘yes’ or ‘no’
Statement
true false
20%

EEG Hypotheses

N400: smaller for expected words e.g., Kutas (1980)

- 300 ms gap: ‘yes’ is more expected than ‘no’ => N400 for ‘no’ vs. ‘yes’
- 1000 ms gap: N400 disappears or even flips?

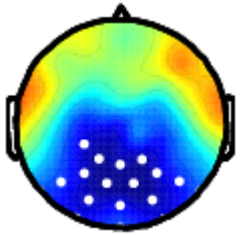
Results: N400



Timing by Response type Interaction for N400

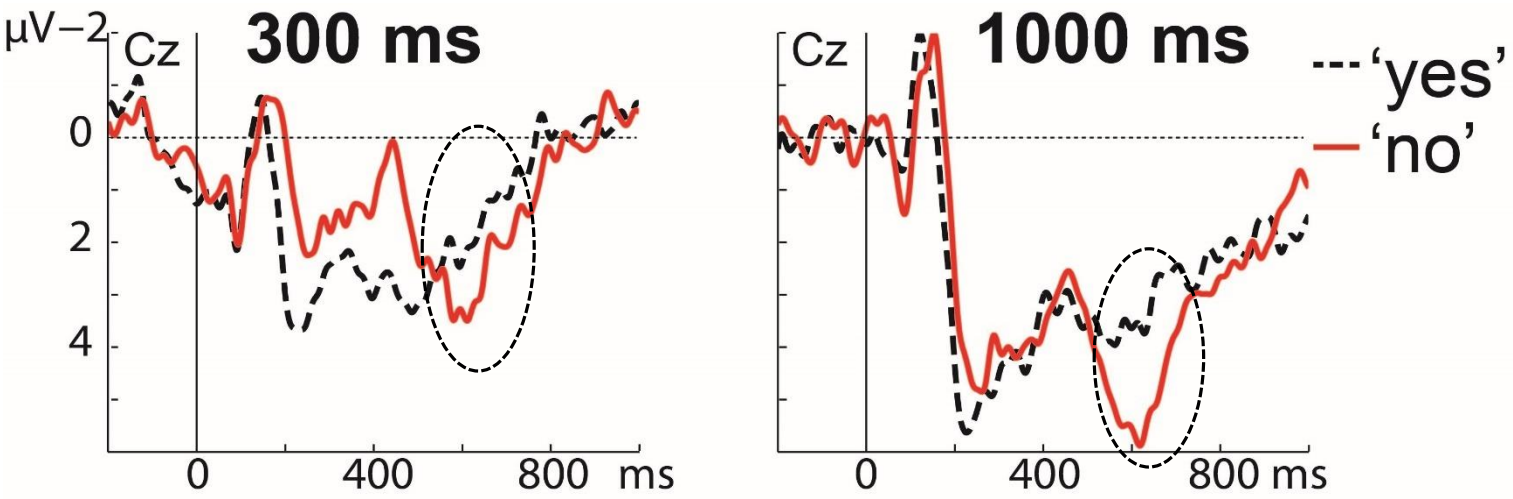
N400 for 'no'

no N400



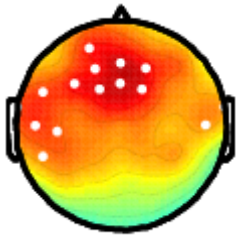
300-500 ms

Results: late positivity



No Timing by Reponse type interaction after 500 ms

Late positivity for 'no' irrespective of timing



500-800 ms

Discussion N400

- After 300 ms gap: larger N400 for ‘no’ => dispreferred is less expected than preferred after a short gap
- After 1000 ms gap: same N400 for ‘yes’ and ‘no’ => expectations converge

$\Rightarrow$ Mere silence changes expectations: expectations for preferred and dispreferred response converge

Why does ‘no’ not become *more* expected?

1. Dispreferreds often qualified or mitigated (e.g., *well*, *maybe*)
   => dispreferred more expected than preferred but not a flat ‘no’
2. General normative expectation for preferred over dispreferred
   + expectation for dispreferred after long gap based on frequencies
   = no effect?

Discussion positivity

- Late positivity for ‘no’ *irrespective* of response timing
- Dispreferred responses are socially accountable actions => require explanations (Garfinkel, 1967; Heritage, 1984)
- A flat ‘no’ might require more analysis or extra processing to understand (e.g., a search for an account)

  cf. positivity to social norm violations (e.g., Leuthold et al., 2015)

$\Rightarrow$ No-responses are socially disaffilliative

Future direction

- Create positive/negative context: when is the gap too long?

References (1)

Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never Say No… How the Brain Interprets the Pregnant Pause in Conversation. *PloS one, 10*(12), e0145474.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. *Scientific Reports, 5*, 12881.

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics, 52*, 46-57.

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 515-535.

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. *Studies in Interactional Sociolinguistics, 13*, 134-184.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. New Jersey: Prentice-Hall, Inc, Englewood Cliffs.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274-279.

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.

Jansen, S., Wesselmeier, H., de Ruiter, J. P., & Mueller, H. M. (2014). Using the readiness potential of button-press and verbal response within spoken language processing. *Journal of Neuroscience Methods, 232*, 24-29.

Kendrick, K. H., & Torreira, F. (2015). The Timing and Construction of Preference: A Quantitative Study. *Discourse Processes, 52*(4), 255-289.

References (2)

Leuthold, H., Kunkel, A., Mackenzie, I. G., & Filik, R. (2015). Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Social Cognitive and Affective Neuroscience, 10*(8), 1021-1029.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223-232.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731.

Local, J., & Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association, 42*(03), 255-280.

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26*, 2530-2539.

Roberts, F., Francis, A. L., & Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of “trouble” in conversation. *Speech Communication, 48*(9), 1079-1093.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696-735.

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: the time course of response planning in conversation. *Frontiers in Psychology, 6*, 284.

Wells, B., & Macfarlane, S. (1998). Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech, 41*(3-4), 265-294.

Thank you!