Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study

Yasuharu Den

Faculty of Letters
Chiba University

LabPhon14@NINJAL

Spontaneous speech

- Ano: sono-ko-wa nihon-ni ki-te: (0.9) ma: (0.3) ano uti: watasi-no it-teru tokoro-wa nihon-go: (0.5) -ga^ (0.3) n nihon-go-gaku-tte-yu-no-o yat-teru tokoro-na-node (0.3) ma: ano: nihon-ni kuru mae-mo moo nihon-go perapera-da-si: (CSJ:D01F0023)

Spontaneous speech

- Ano: sono-ko-wa nihon-ni ki-te: (0.9) ma: (0.3) ano uti: watasi-no it-teru tokoro-wa nihon-go: (0.5) -ga^ (0.3) n nihon-go-gaku-tte-yu-no-o yat-teru tokoro-na-node (0.3) ma: ano: nihon-ni kuru mae-mo moo nihon-go perapera-da-si: (CSJ:D01F0023)
- Various sorts of disfluencies:
  - Fillers, suspensions, repairs, repetitions, & segment lengthening

Today's topic

- Vowel lengthening (aka prolongation)
  - Non-lexical stretching of speech segments
  - Occurring everywhere in an utterance
- Examples from the previous excerpt
  - End of fillers: *ano:*, *ma:*
  - End of phrases: *uti:*, *nihon-go:*
  - End of clauses: *ki-te:*, *perapera-da-si:*
  - Over 90% of lengthening in Japanese occurs at the end of words (Den, 2003).
- Q: What factors behind vowel lengthening?

Outline

1. Background
2. Data, annotation, and statistical methods
3. Analysis 1: Lengthening at the beginning of utterances: Fillers and conjunctions
4. Analysis 2: Lengthening at the beginning of utterances: *Wa*-marked phrases
5. Analysis 3: Lengthening at the end of utterances
6. Discussion

BACKGROUND

Vowel lengthening

- Vowel lengthening has been studied in phonology, phonetics, and speech synthesis research.
- Various factors
  - Final lengthening at various levels (Klatt, 1975)
    - Word < Accentual phrase < Intonation phrase
  - Pre-pausal lengthening
  - Compensatory effect of the surrounding phonemes, e.g. mora-timed rhythm in Japanese
  - Simultaneous lengthening of successive phonemes within a syllable (Campbell & Isard, 1991)

Factors used in speech synthesis

- These factors, among others, have been applied to speech synthesis (Kaiki et al., 1990):
  - compensatory effect of the surrounding consonants
  - position of the vowel in the word, the prosodic phrase, and the utterance
  - presence of the following pause
  - syntactic category of the word
  - inherent duration of the vowel
  - overall speech rate of the speaker

Prolongation in spontaneous speech

- Swedish (Eklund, 2001)
  - focused on phonological & morpho-syntactic factors such as phone type, position in the word, lexical factors, and word class
- Japanese (Den, 2003)
  - found some strategies in prolonging speech segment used by Japanese speakers
- Mandarin (Lee et al., 2004)
  - took functional difference into account such as hesitation, emphasis, and feedback

Possible other factors

- Only linguistic factors so far
- In spontaneous speech, other factors may affect vowel lengthening.
- Planning load = Cognitive factor
  - In spontaneous discourse, speech planning is done on the fly.
  - Speakers may take extra time in planning complex utterances.
  - On these occasions, they signal their problems in the form of disfluencies (Clark, 2002):
    - Fillers, repetitions, repairs, and segment lengthening

Previous findings in my studies

- Utterance initial fillers and conjunctions
  - Filler *e*: positively correlated with the duration of the utterance (under some conditions) (Watanabe & Den, 2010)
  - Conjunction *de*: no such correlation (Den, 2009; Watanabe & Den, 2010).
- Utterance initial *wa*-marked topic phrases
  - *Wa*: positively correlated with the duration of the rest of the utterance (Watanabe & Den, 2010; Den & Nakagawa 2013).
- End of clauses
  - Final mora: positively correlated with the duration of the next clause (within an utterance) (Koiso & Den, 2013).

Problems of the previous studies

- Phonological and morpho-syntactic factors were not fully considered (nor controlled).
- The cause-effect relationship, i.e. which is dep. variable and which is indep. variable, was not consistent across studies.
- The relationship among lengthening at difference places was not investigated.

- In this talk, I solves some of these problems.

DATA, ANNOTATION, AND STATISTICAL METHODS

Agenda for methodology

- The study of spontaneous speech
  - Difficult to apply experimental methods
  - Important to investigate natural speech data
- But, natural speech data is messy.
- Requirements for dealing with natural speech data
  - Big amount of data

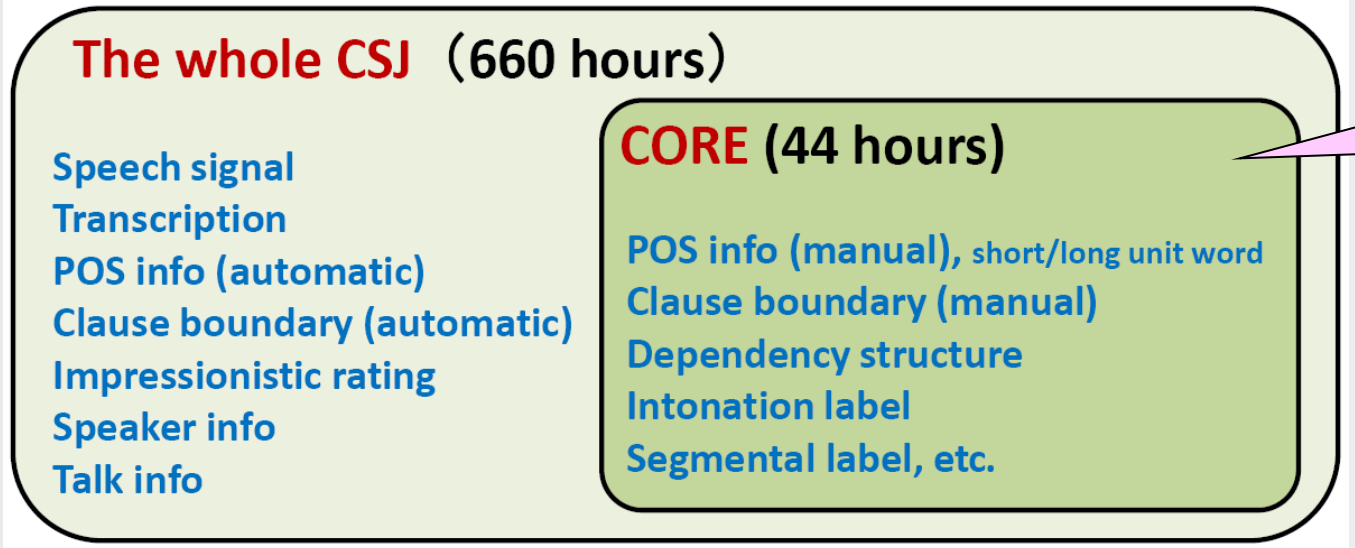
    -> *Corpus of Spontaneous Japanese*
  - Control of confounding variables

    -> Data selection & covariates
  - Adequate statistical method

    -> Mixed-effects model

Data

- *Corpus of Spontaneous Japanese* (CSJ)
  - Large-scale corpus of spontaneous Japanese, developed by NINJAL
  - Mostly monologs (625 hours)
    - Academic presentations and speech on everyday topics



Detailed annotation

Data analyzed
107 talks
20 hours
9.8K clause units
230K words

Annotation

- Linguistic annotations at various levels
  - Phonemes
    - Starting and ending times, their uncertainty, devoicing, etc.
  - (Long- & short-unit) words
    - Part-of-speech, conjugation form, dictionary form, etc.
  - Accentual phrases
    - Break indices and boundary tones (based on X-JToBI)
  - Bunsetsu phrases
    - Dependency structures
  - Clause units (regarded as utterances)
    - Clause boundary (CB) types
- Compiled in a relational database (Koiso et al., 2014)

CB types

- Absolute (AB)
  - sentence ending in usual sense

  E.g. Tokyo-ni iki-masu
  Tokyo-DAT go-POL
  I will go to Tokyo.
- Strong (SB)
  - clause boundary with coordinate particle

  E.g. Tokyo-ni iki-masu-ga
  Tokyo-DAT go-POL-but
  I will go to Tokyo, but …
- It is sometimes better to consider other types of clauses and phrases as independent utterances.
  - Weak (WB)

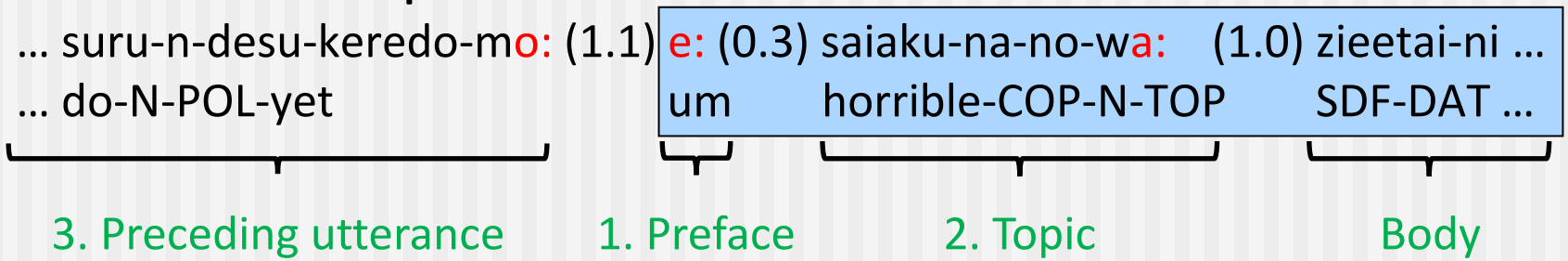
  E.g. Tokyo-ni iku-node
  Tokyo-DAT go-because
  Because I go to Tokyo.
  - Non-Clausal (NCB)

  E.g. Tokyo-ni
  Tokyo-DAT
  To Tokyo

AB > SB > WB > NCB

Variables

- Schematic representation of the utterance



  - Preface and topic may be missing.
- Dependent variable
  - Duration of the final vowel
- Independent variables
  - CB type of the preceding utterance
  - Duration of the body

Degree of disjuncture between utterances

Complexity of the utterance

Data selection

- Use only reliable data
  - E.g. exclude cases where the phoneme boundaries are uncertain
- Use only major categories
  - E.g. focus on frequent preface items such as filler *e* and conjunction *de*
- Use only simple cases
  - E.g. focus on topics with the simplest structure, i.e. Noun/Pronoun-*wa*

Statistical analysis

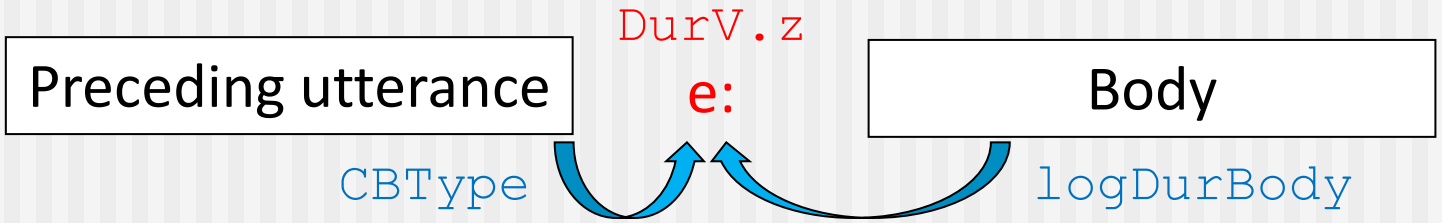
- To consider the inherent duration of the vowel and the overall speech rate of the speaker
  - z-score transformation (cf. Campbell & Isard, 1991)

$$Dur_z = \frac{\log(Dur_{raw}) - mean}{sd}$$

    - mean/sd: calculated for each phoneme and each speaker
- To consider the hierarchical structure of the corpus data, i.e. clusters according to speakers
  - mixed-effects models (Baayen, 2008)
  - random intercept for speakers and word forms

ANALYSIS 1:
FILLERS AND CONJUNCTIONS

Method



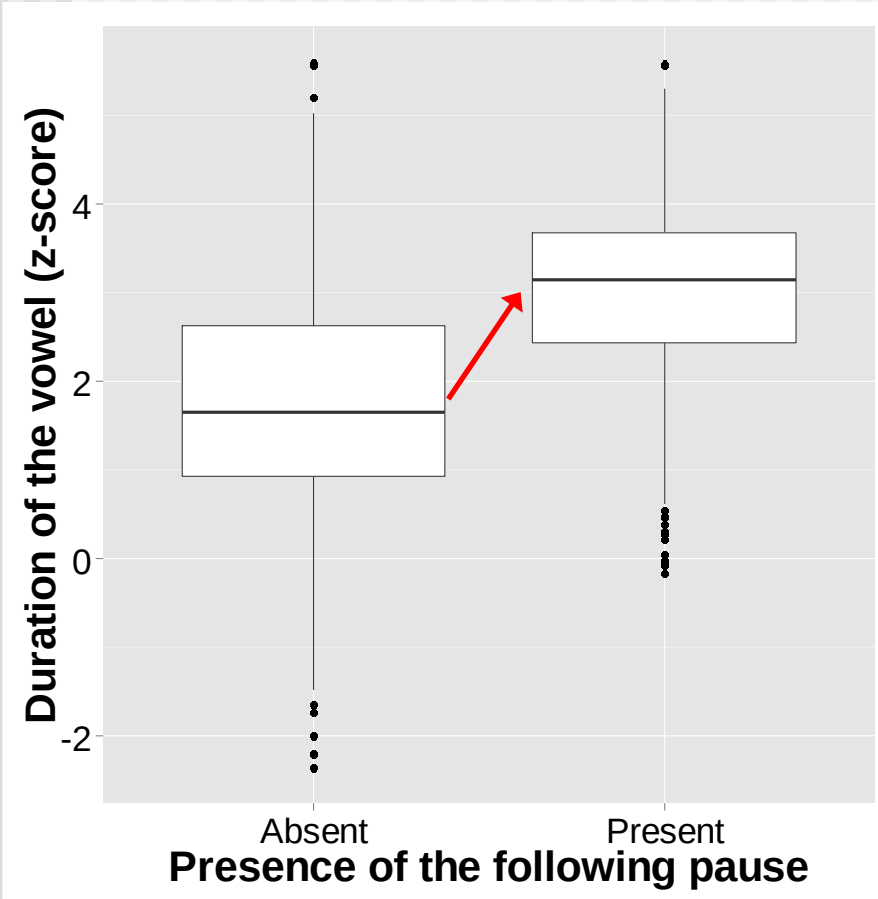
- Data selection
  1. Limited to four major categories (> 66%)
     - Filler *e*, Filler *ma*, Filler *ano*, & Conjunction *de*
  2. Excluding uncertain phoneme boundaries and non-canonical pronunciation
  3. Limited to simple patterns, i.e. utterance initial fillers/conjunctions followed by no other preface items
- Data analyzed

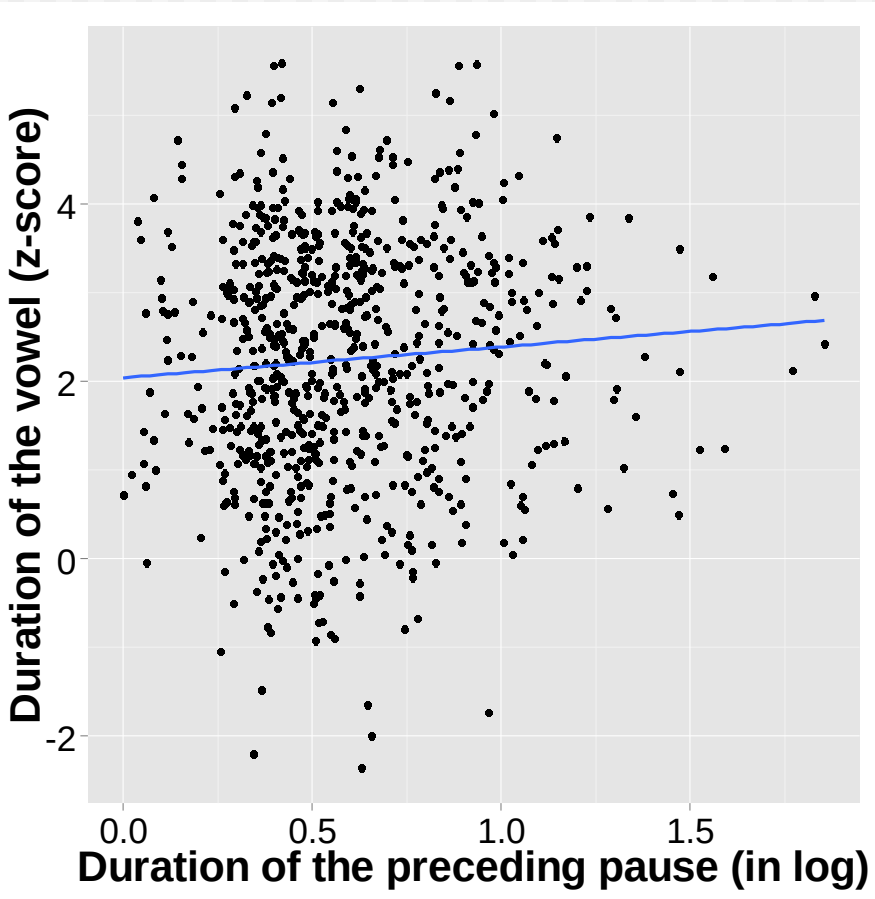
Filler *e*	Filler *ma*	Filler *ano*	Conj *de*
761	615	353	839

Method (cnt'd)

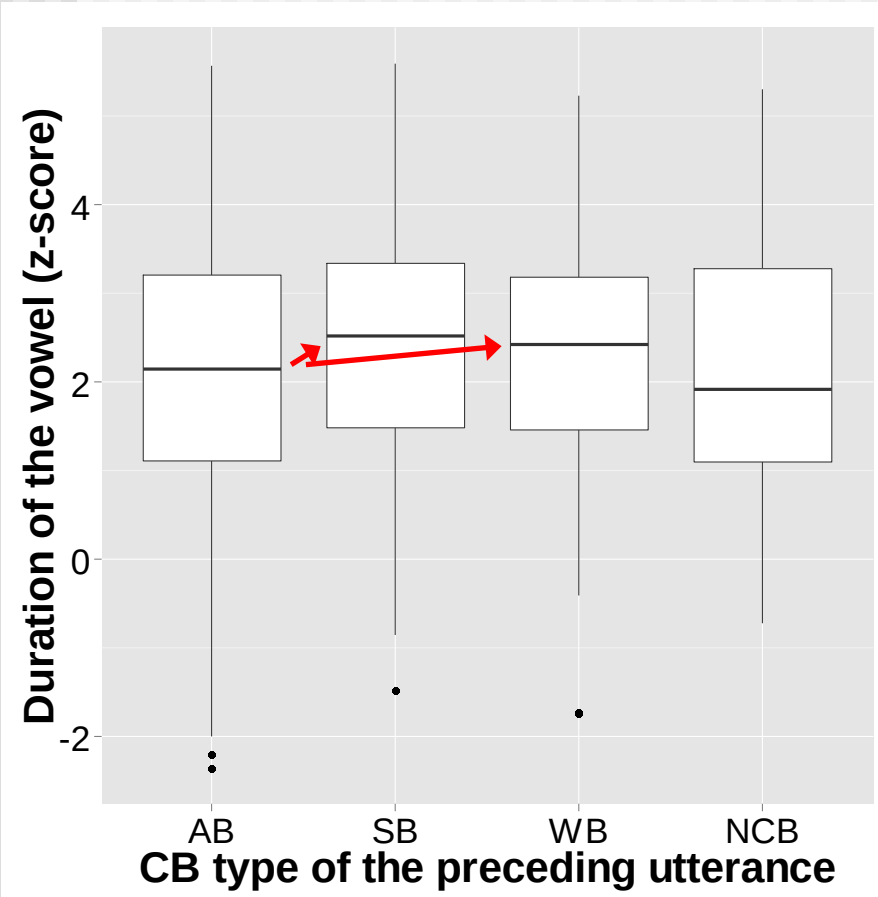
- Variables
  - Independent variables
    - CB type of the preceding clause (`CBType`)
    - Duration of the body (in log) (`logDurBody`)
    - Their interaction (not significant, removed)
  - Covariates
    - Duration of the preceding consonant (`DurC.z`)
    - Presence of the following pause (`ifFolPause`)
    - Duration of the preceding pause (in log) (`logDurPrePause`)
    - Presence of the topic (`ifTopic`)
  - Random effects
    - Intercept for speakers
- Parameter estimation
  - Maximal Likelihood Estimation by `lmer` of R language
  - *p*-values obtained by likelihood ratio tests

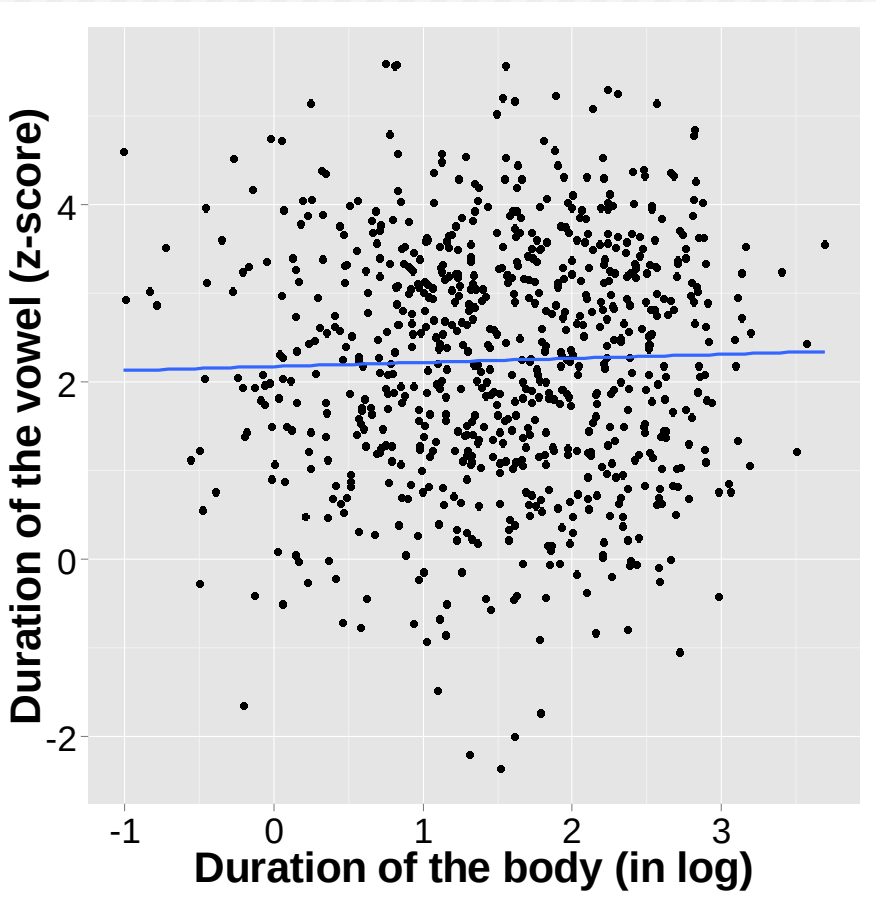
Results: *e* vs covariates





Results: *e* vs indep. variables





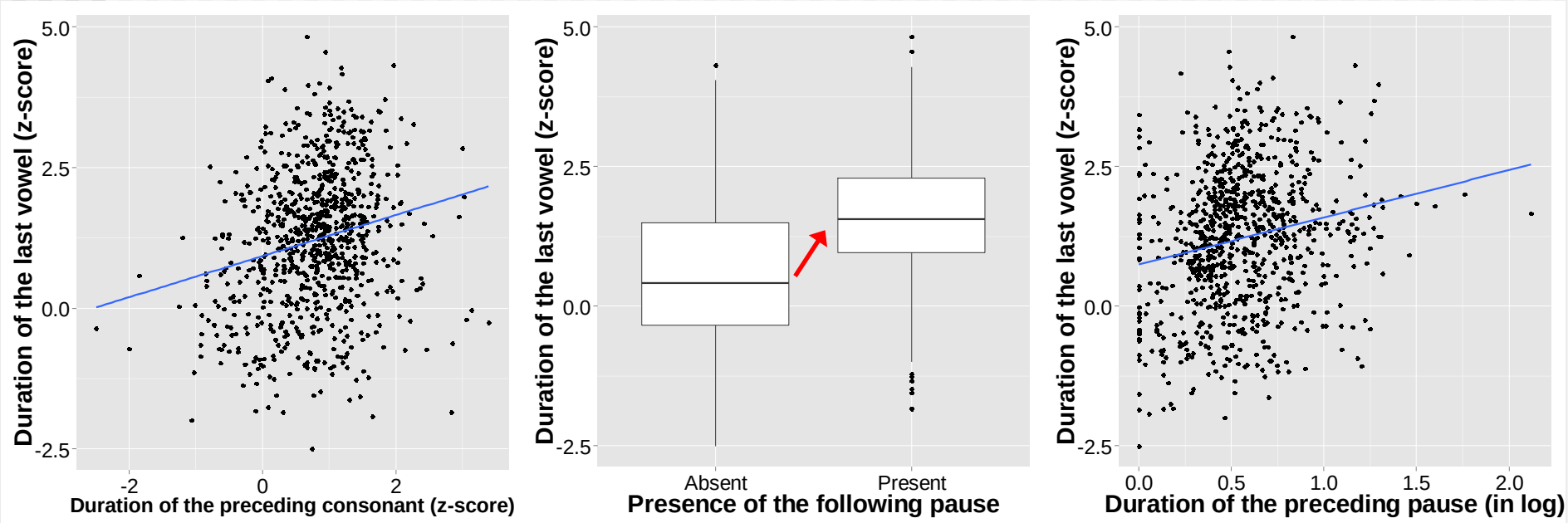
Results: Mixed-effects model

- All covariates had a significant effect.
- The CB type had a significant effect (AB < SB, WB), but the body duration did not.

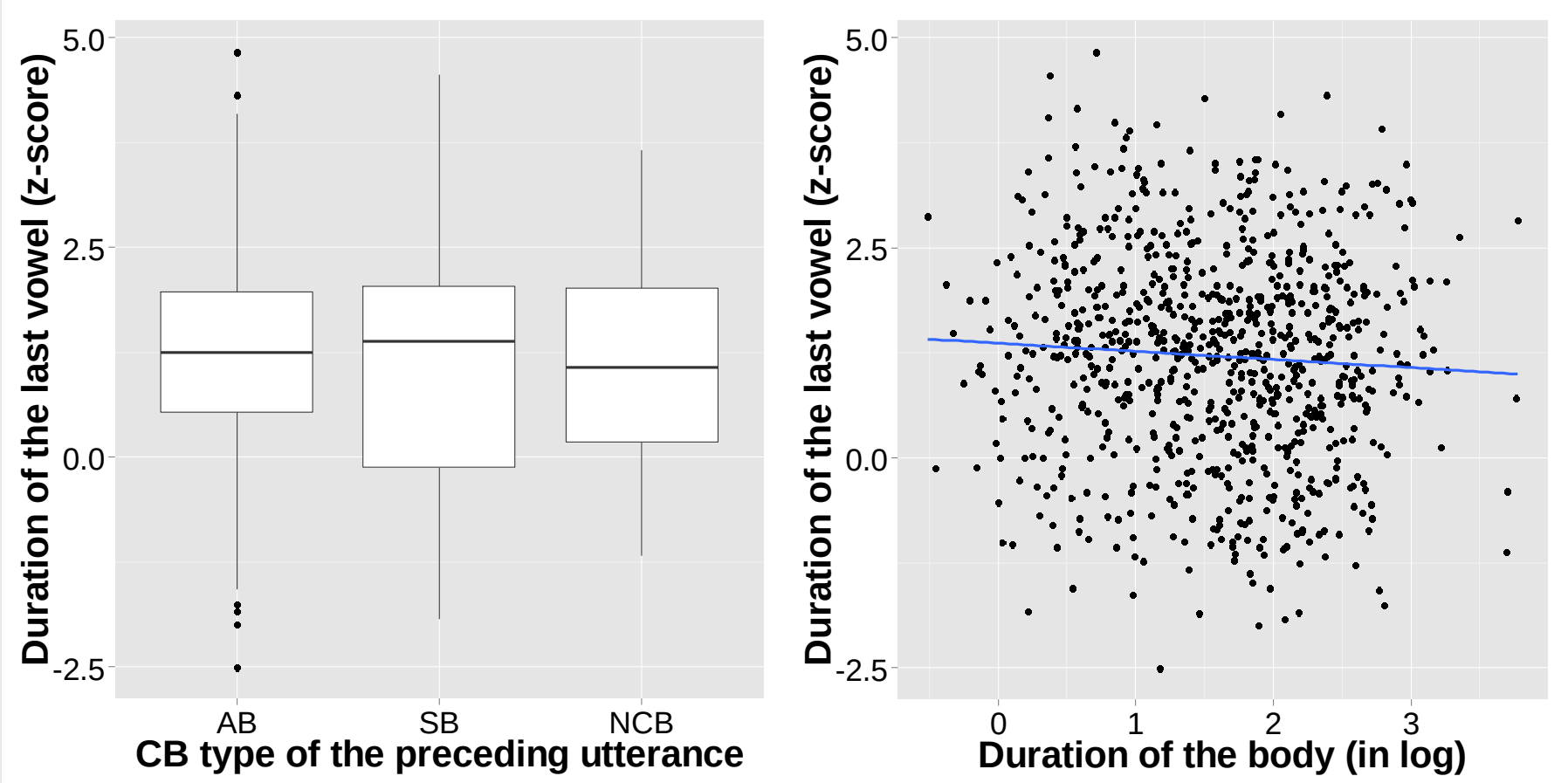
	Coef.	SE	*t* value	*p* value
CBType=SB	.356	.100	3.55	
CBType=WB	.332	.137	2.43	.002
CBType=NCB	.053	.174	.31	
logDurBody	.066	.049	1.35	.177
ifFolPause	1.214	.087	13.88	.000
logDurPrecPause	.352	.174	2.03	.043
ifTopic	-.150	.106	-1.42	.157

$\sigma = 1.07$, $\sigma_{Speaker} = .65$

Results: *de* vs covariates



Results: *de* vs indep. variables



Results: Mixed-effects model

- All covariates had significant effects.
- No significant effects of the CB type or the body duration

	Coef.	SE	*t* value	*p* value
CBType=SB	-.055	.088	-.62	.804
CBType=NCB	.028	.209	.13	
logDurBody	-.065	.046	-1.43	.154
DurC.z	.345	.051	6.72	.000
ifFolPause	.950	.076	12.47	.000
logDurPrecPause	.753	.142	5.30	.000
ifTopic	-.173	.093	-1.88	.062

$\sigma = .98, \sigma_{Speaker} = .47$

Summary of the results

- All phonological covariates had significant effects.
- The effect of the CB type was significant in fillers *e* and *ma*, but the effect of the body duration was significant only in filler *ma*.

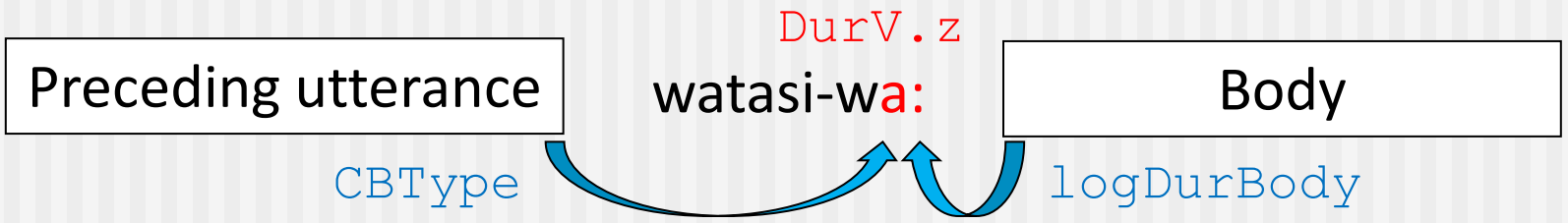
	Filler *e*	Filler *ma*	Filler *ano*	Conj *de*
CBType	AB < SB, WB	AB < SB, WB	ns	ns
logDurBody	ns	+	ns	ns
DurC.z		+	+	+
ifFolPause	+	+	+	+
logDurPrecPause	+	+	ns	+
ifTopic	ns	ns	ns	ns

Summary of Analysis 1

- Lengthening of the last vowel in utterance initial fillers and conjunctions is
  - consistently affected by phonological factors:
    - the duration of the preceding consonant
    - the presence of the following pause
  - but not always affected by cognitive factors
- Among fillers, *ma* is most affected by cognitive factors, *e* next, and *ano* least.
- Conjunction *de* seems independent of cognitive factors.

ANALYSIS 2:
WA-MARKED PHRASES

Method



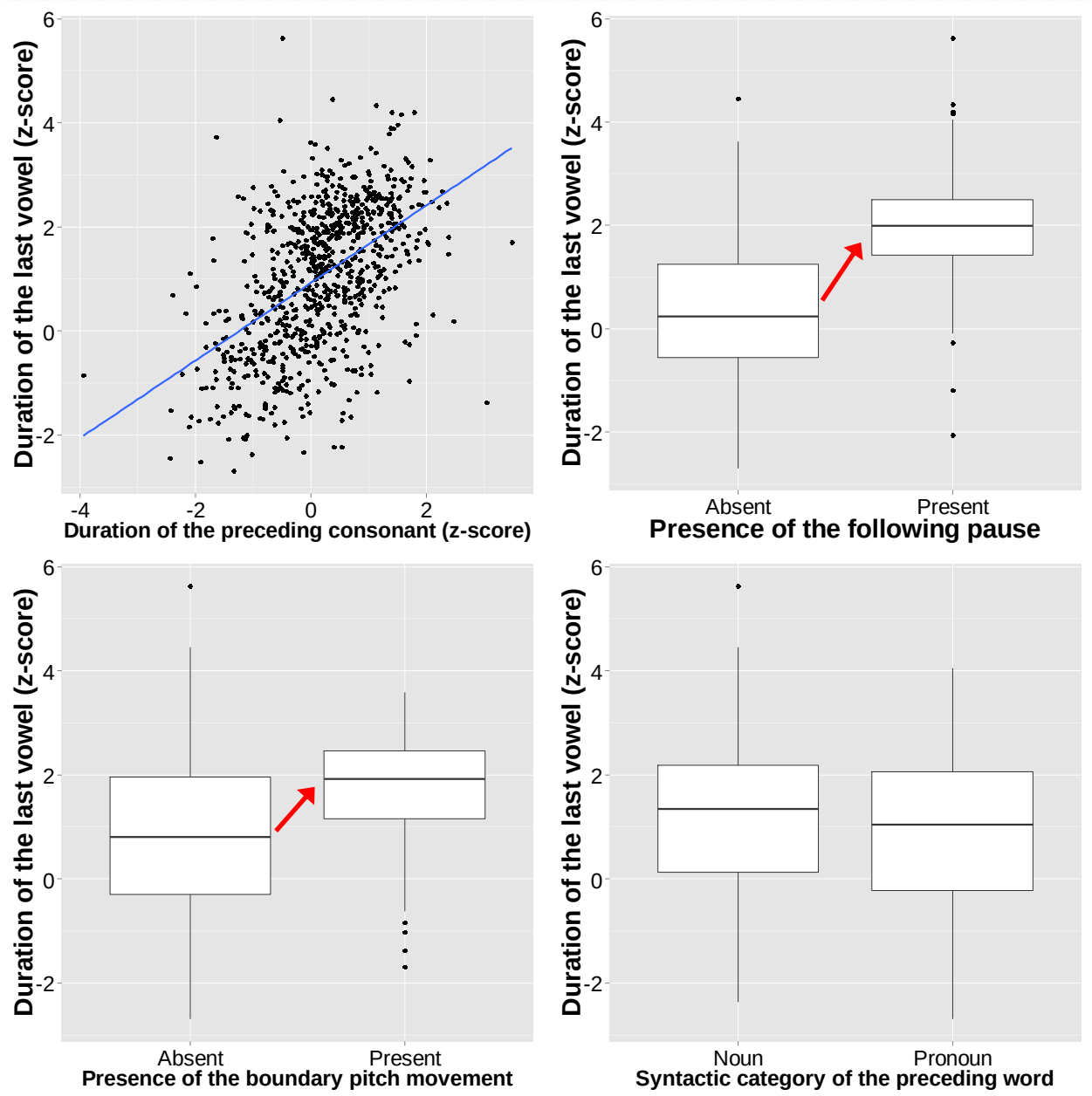
- Data selection
  1. Limited to those starting with noun or pronoun (> 97%)
  2. Excluding uncertain phoneme boundaries and non-canonical pronunciation
  3. Limited to simple phrases, i.e. Noun/Pronoun-*wa* (including Noun/Pronoun-*toiu-no-wa* and the like)
- Data analyzed

Noun-*wa*	Pronoun-*wa*
464	337

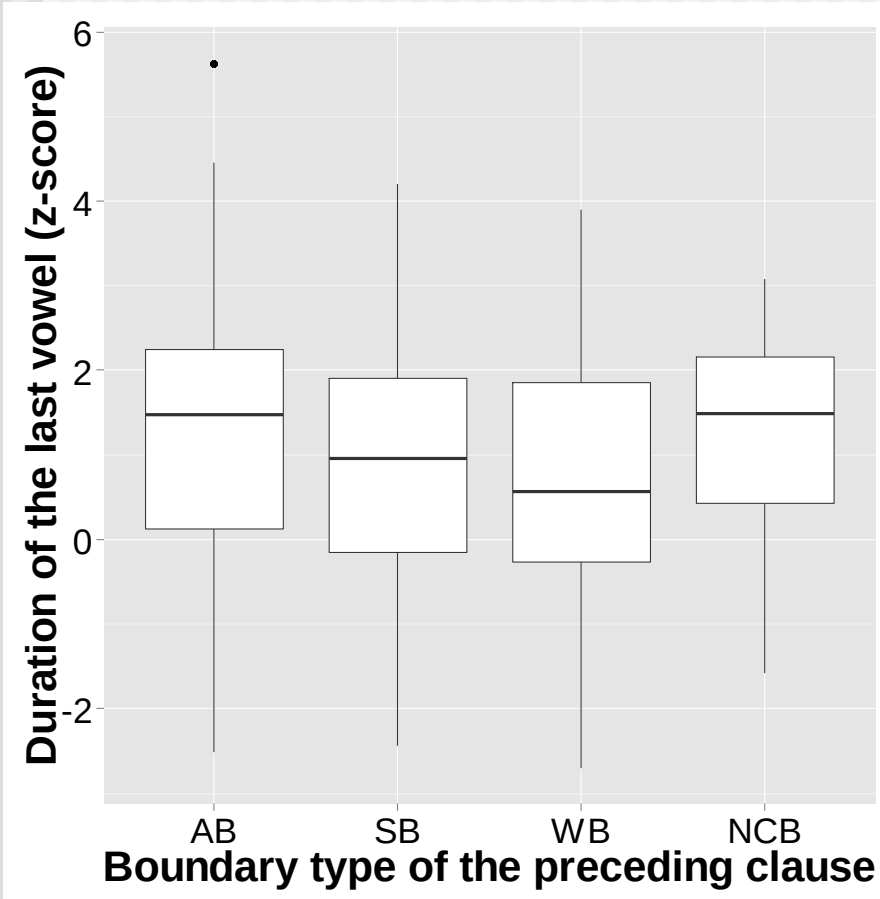
Method (cnt'd)

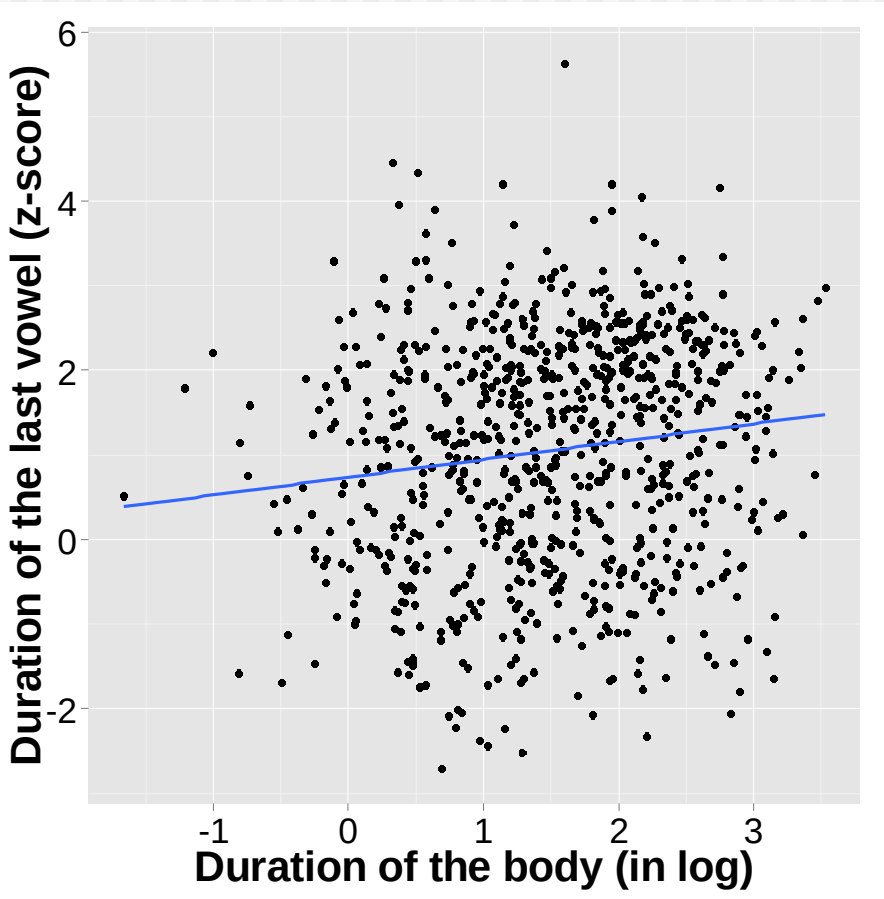
- Variables
  - Independent variables
    - CB type of the preceding clause (`CBType`)
    - Duration of the body (in log) (`logDurBody`)
    - Their interaction (not significant, removed)
  - Covariates
    - Duration of the preceding consonant (`DurC.z`)
    - Presence of the following pause (`ifFolPause`)
    - Presence of the boundary pitch movement (`ifBPM`)
    - Syntactic category of the preceding word (`ifPrePronoun`)
    - Presence of the preface (`ifPreface`)
  - Random effects
    - Intercept for speakers
    - Intercept for word forms (not significant, removed)

Results: *Wa* vs covariates



Results: *Wa* vs indep. variables





Results: Mixed-effects model

- All phonological covariates had significant effects.
- The effect of the body duration was also significant.

	Coef.	SE	*t* value	*p* value
CBType=SB	-.172	.083	-2.08	
CBType=WB	-.270	.130	-2.08	.079
CBType=NCB	-.028	.148	-.19	
logDurBody	.092	.041	2.23	.026
DurC.z	.472	.042	11.34	.000
ifFolPause	1.148	.079	14.54	.000
ifBPM	.346	.091	3.80	.000
ifPrePronoun	-.065	.075	-.88	.381
ifPreface	.037	.077	.48	.629

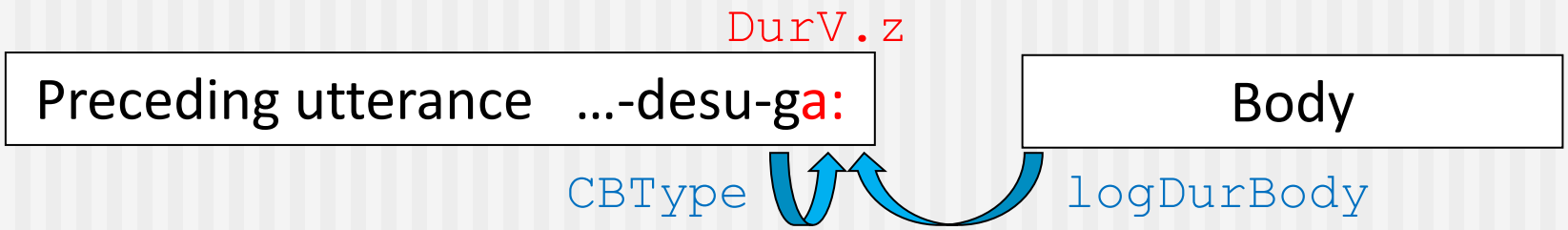
$\sigma = .97$, $\sigma_{\text{Speaker}} = .31$

Summary of Analysis 2

- Lengthening of the vowel *a* of *wa* in utterance initial topic phrases is
  - affected by phonological factors:
    - the duration of the preceding consonant
    - the presence of the following pause
    - the presence of the boundary pitch movement
  - and also affected by some cognitive factor, i.e. the complexity of (the rest of) the utterance

ANALYSIS 3:
END OF UTTERANCE

Method



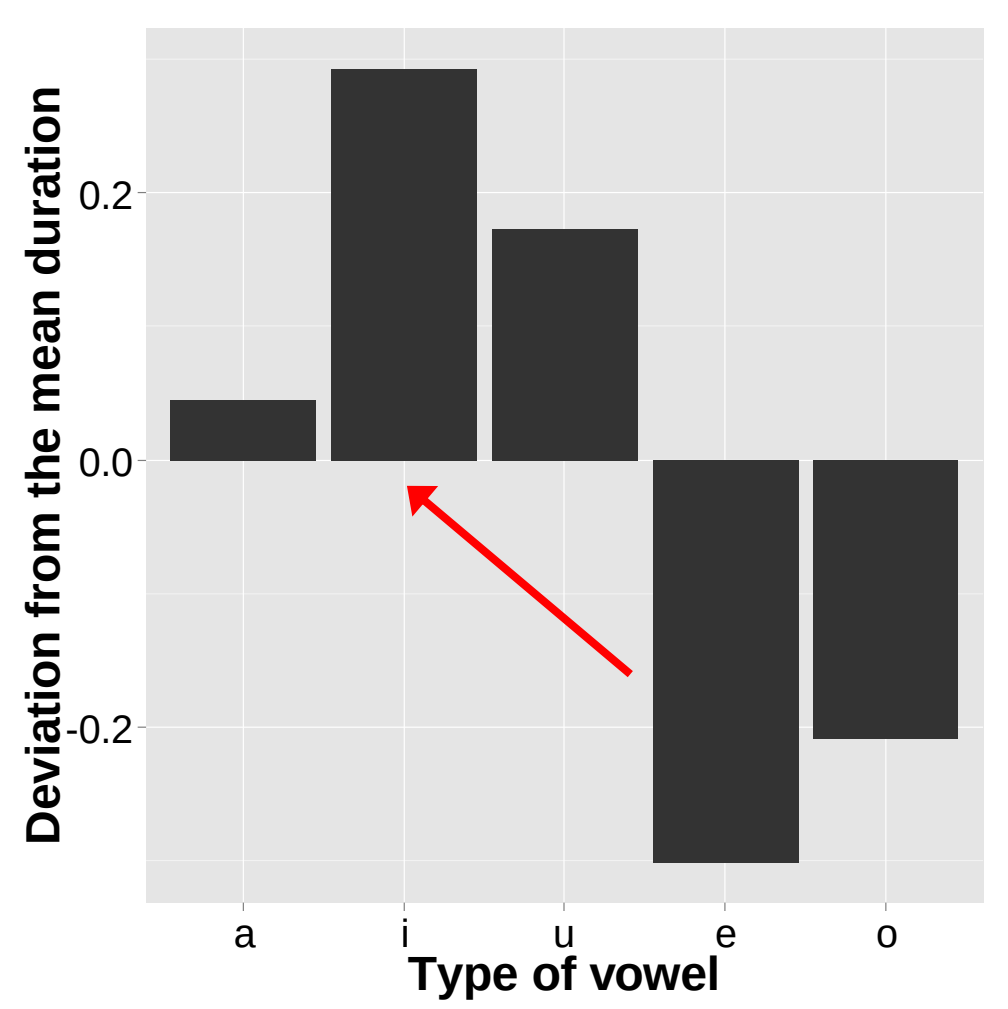
- Data selection
  1. Limited to those ending with particle or auxiliary verb (> 94%)
  2. Excluding uncertain phoneme boundaries, non-canonical pronunciation, and devoiced vowels
  3. Limited to those coincident with AP boundary
- Data analyzed

AB	SB	WB	NCB
2005	2940	738	253

Method (cnt'd)

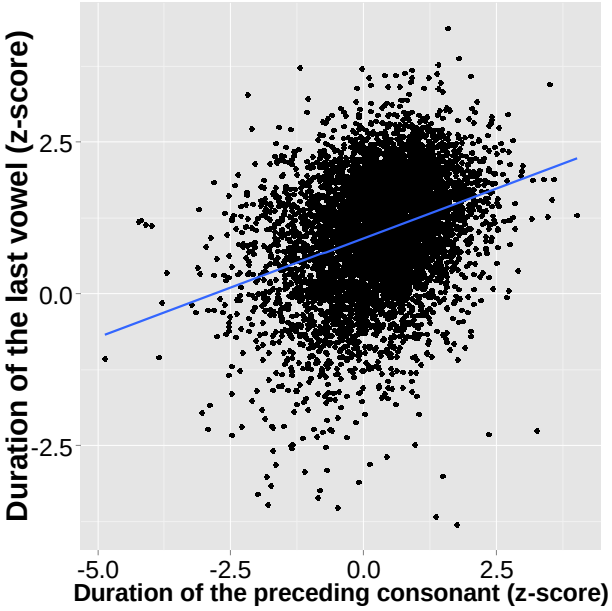
- Variables
  - Independent variables
    - CB type of the utterance (`CBType`)
    - Duration of the body (in log) (`logDurBody`)
    - Their interaction
  - Covariates
    - Type of the vowel (`VEntity`)
    - Duration of the preceding consonant (`DurC.z`)
    - Presence of the following pause (`ifFolPause`)
    - Presence of the boundary pitch movement (`ifBPM`)
    - Syntactic category of the word (`ifAuxV`)
    - Presence of the preface (`ifPreface`)
    - Presence of the topic (`ifTopic`)
  - Random effects
    - Intercept for speakers
    - Intercept for word form

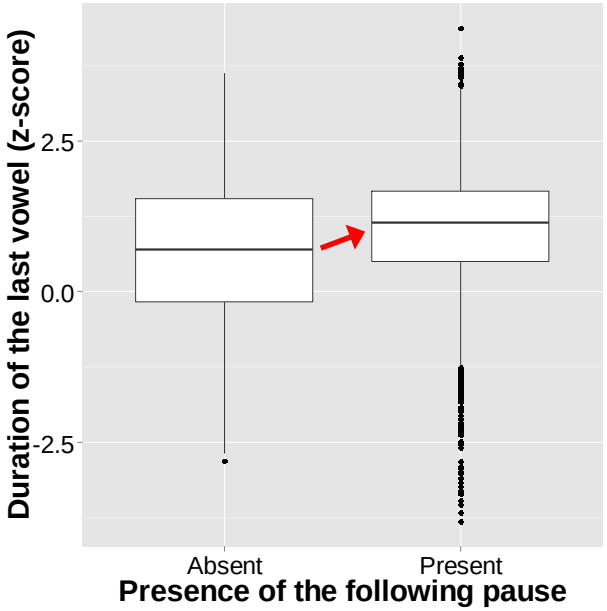
Results: End of utt. vs vowel type

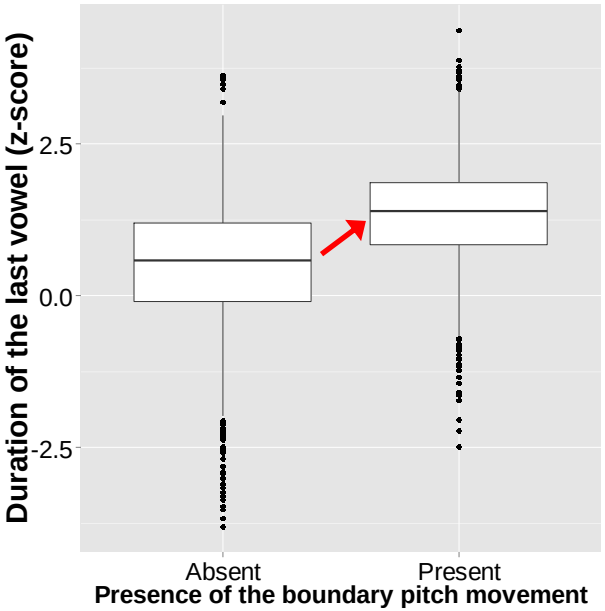


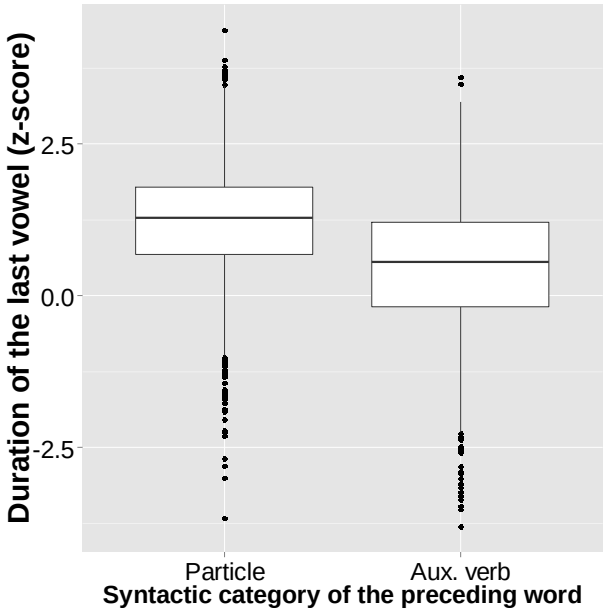
- *a, i, & u* were longer than *e* & *o* at the end of utterance (on z-score scale).

Results: End of utt. vs covariates

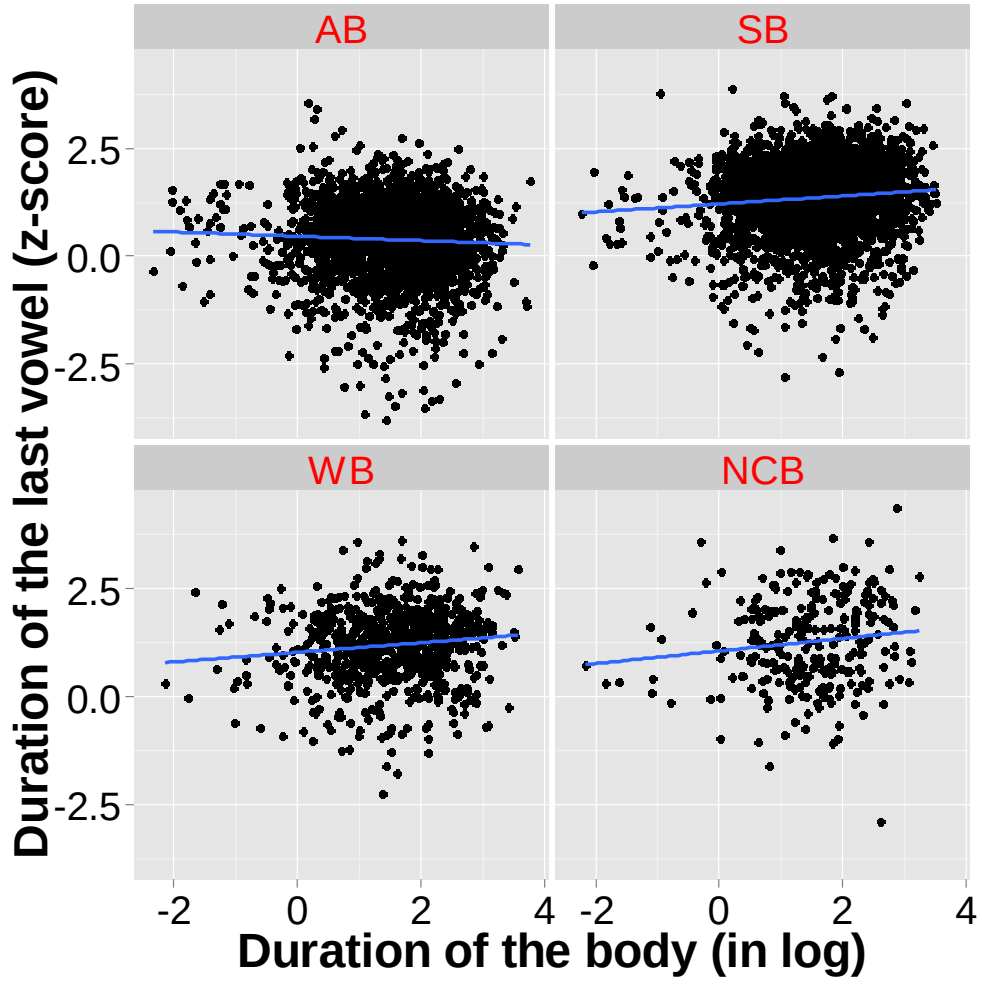








Results: Mixed-effects model

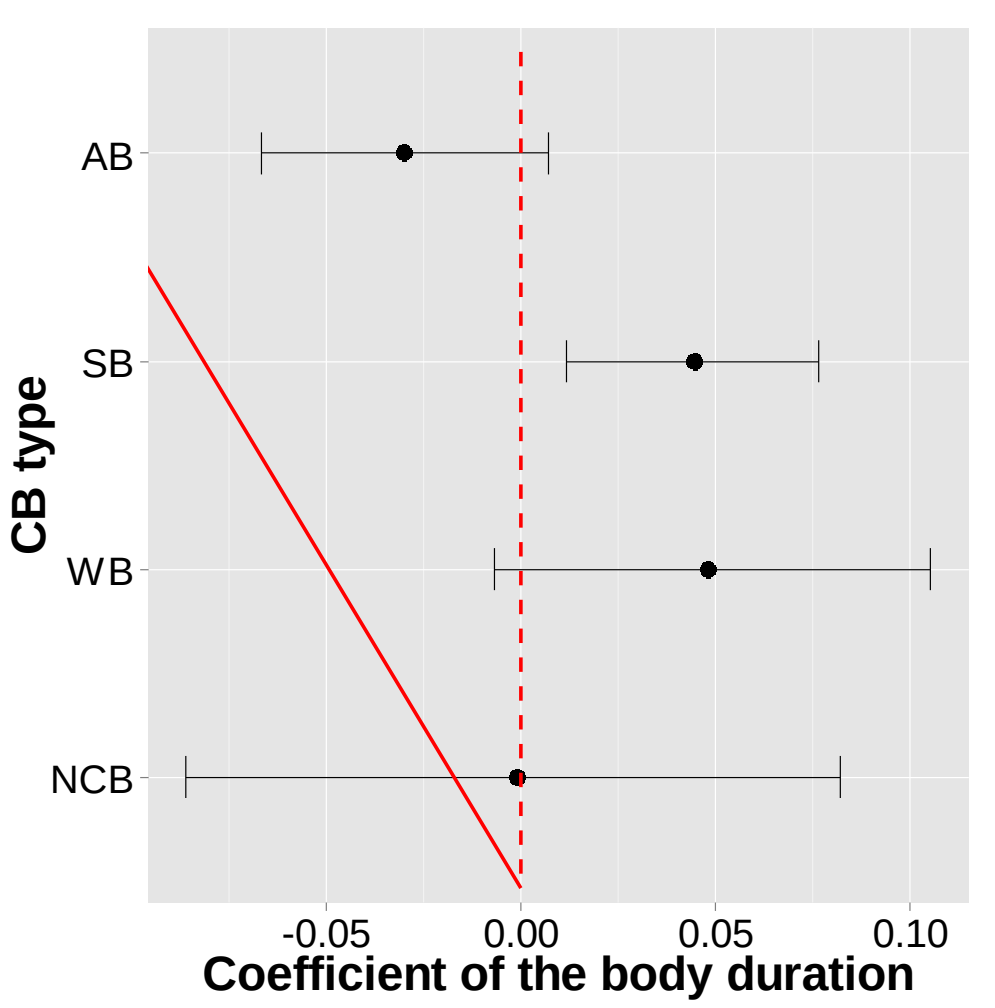


- All phonological covariates had significant effects.
- Significant interaction between the CB type and the body duration (p = .002 by LRT)

	Coef.	SE	*t* value	*p* value
VEntity				.016
DurC.z	.112	.012	9.60	.000
ifFolPause	.460	.036	12.95	.000
ifBPM	.553	.026	21.44	.000
ifAuxV	-.082	.084	-.98	.335
ifPreface	.019	.021	.92	.360
ifTopic	-.018	.026	-.70	.484

σ = .72, $\sigma_{Speaker}$ = .23, σ_{Orth} = .23

Results: Coefs. of body duration



- To obtain precise estimates of the coefficients of the body duration for each CB type, we applied MCMC technique using JAGS language (Kruschke, 2011)
- The body duration had a significant positive coefficient only when the CB type was the strong boundary.

Summary of Analysis 3

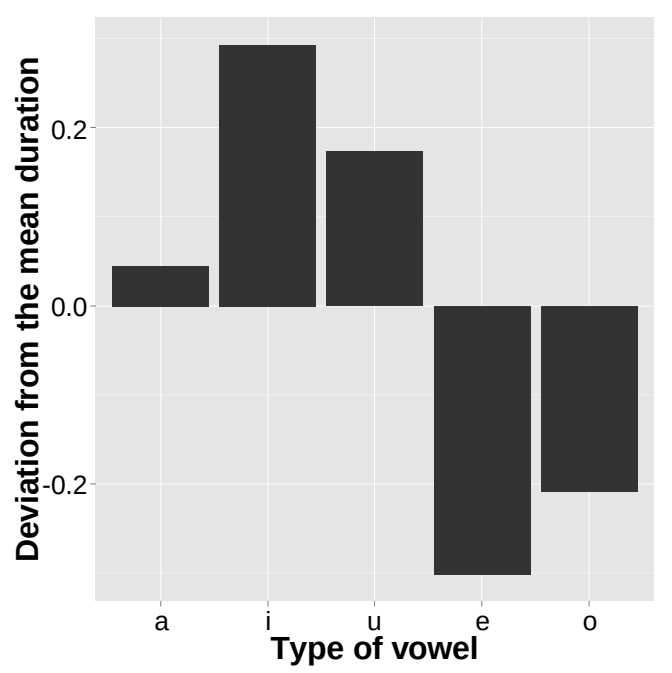
- Lengthening of the last vowels of utterances is
  - affected by phonological factors:
    - the duration of the preceding consonant
    - the presence of the following pause
    - the presence of the boundary pitch movement
    - the type of the vowel
  - and also affected by the complexity of the following utterance under some conditions, i.e. when ending with a strong boundary

DISCUSSION

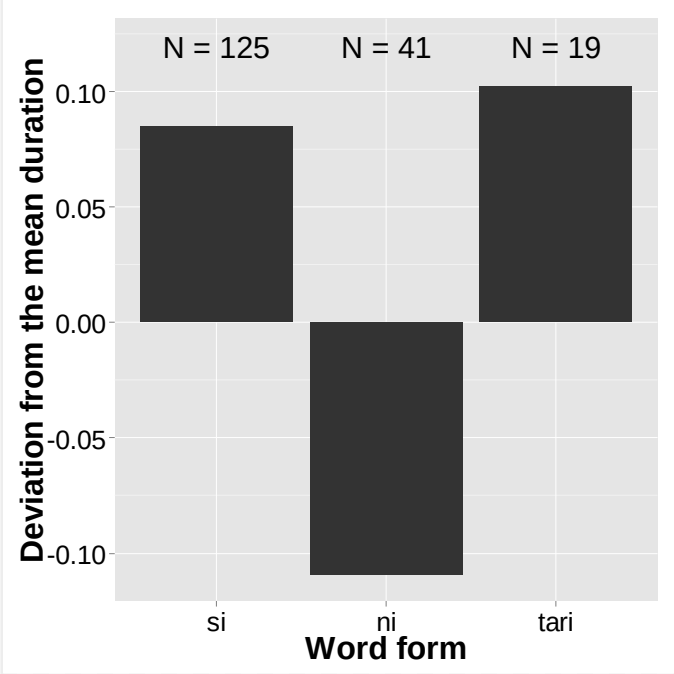
Phonological factors

- At all places, our phonological factors have reliable effects, i.e. vowel lengthening is enhanced when
  - the duration of the preceding consonant is longer;
  - the vowel is followed by a pause; and
  - the vowel bears boundary pitch movement
- The effect of the preceding consonant is supplementary rather than compensatory, suggesting that the entire mora, not just the vowel, is lengthened.

Effect of vowel type



- The degree of lengthening is affected by the vowel type.
  - *i* has the strongest effect, although its inherent duration is short.



- This effect is attributed mainly to a particular lexical item.
  - coordinate particle *si*, which appears at strong boundaries

Complexity effect

- The duration of the utterance body, i.e. the complexity, sometimes affects vowel lengthening.

	Filler *ma*	Topic marker *wa*	Last mora at SB
Coef. of `logDurBody`	.157	.092	.048

- The significant effect found at the last mora of the preceding utterance with a strong boundary may suggest that Japanese speakers use early signal for upcoming troubles.
- But, these coefficients are rather small compared with those of the covariates, e.g. 1.053 for the following pause and .342 for BPM in the case of *wa*.

Disjuncture effect

- The CB type, i.e. the degree of disjuncture between utterances, is sometimes relevant.
  - Filler *e* is longer at SB & WB than at AB.
  - The complexity effect on the last vowel of the preceding utterance is significant only when the preceding utterance ends with SB.
- Two possible explanation for weaker effect at AB
  1. The data for AB is distorted.
     - In analysis 3, the data selection step removed 50% of the data for AB (due to devoicing in *desu* & *masu*) but only 15-25% for SB & WB.
  2. Some different cognitive process is involved at AB.
     - E.g. discourse-level planning

Summary

- Vowel lengthening in spontaneous Japanese
  - Phonological factors
  - Cognitive factors
- Further Q: Relationship among lengthening at different places?
  - Complementary, supplementary, or independent?
  - Related to different functions?
- Ready to go out of laboratories!
  - Adequate corpora and analytic methods

Thank you for your kind attention

and

Let’s enjoy fireworks!

Special thanks to:

Michiko Watanabe

Natsuko Nakagawa

and Hanae Koiso