

Linking dynamic speaker variation to real time sound change: perspectives and issues

Florent Chevalier

Université de Poitiers, University of Glasgow

GULP Lablunch, 10 May 2018



Introduction

Ongoing PhD (2017-2020):

- **De la convergence inter-locuteur au changement phonétique** : accommodation linguistique et changement phonétique dans l'anglais parlé à Glasgow
- Jointly supervised by Sylvie Hanote (Poitiers) and Jane Stuart-Smith (Glasgow)
- Funded by the French Ministry for Higher Education, Research and Innovation



Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION

Introduction

Today's menu:

- Starter: a taste of this PhD's *raison d'être*
- Main: jumping into data with the first pilot
- Dessert: “How’s your PhD going?”
- Digestif: problem-solving session (BYOB)

Background

- When looking at sound change in progress, we only use aggregates for each speaker (Cukor-Avila & Bailey 2013).
- However, we know about short term (Pardo 2012) and medium term (Sonderegger 2012) change.
- Short term accommodation (see Giles 1979) is thought to be the starting point of long term community level change (Trudgill 1986).

Research questions

Three main questions :

- How does accommodation within speakers in the **short term** relate to change in the **long term**?
- How is speech accommodation constrained by **linguistic, social** and **situational** factors?
- How did vowel **quality** and **quantity** change in Glaswegian English across the 20th century?



SOUNDS OF THE CITY



Fine phonetic variation and sound change: A real-time study of Glaswegian

<http://soundsofthecity.arts.gla.ac.uk/>

Oct 2011-Dec 2014



The Leverhulme Trust

Method

- The sample will include all age groups (all 3 generations for all 4 decades of recording), different kinds of interactions (interviews, discussions between peers ...), same/mixed-gender interactions.
- Evolution of vowel quality and quantity will be monitored dynamically (within interactions) and in real time (over decades).

First pilot: presentation

- Checking if raw data suggest a pattern in variation within an interaction
- Finding out the most efficient way to extract clean data/exhaustive information from the corpus
- Deciding of a closed list of tokens to keep/exclude
- Identifying which processes can be automated in the near future
- Testing the accuracy of the corpus' automated segmentation

First pilot: presentation

- One interaction: two men from the 90M group
- One vowel: /i/ (all stressed instances, N = 984 and then down to 413)
- One feature: temporal, i.e. vowel length (SVLR)

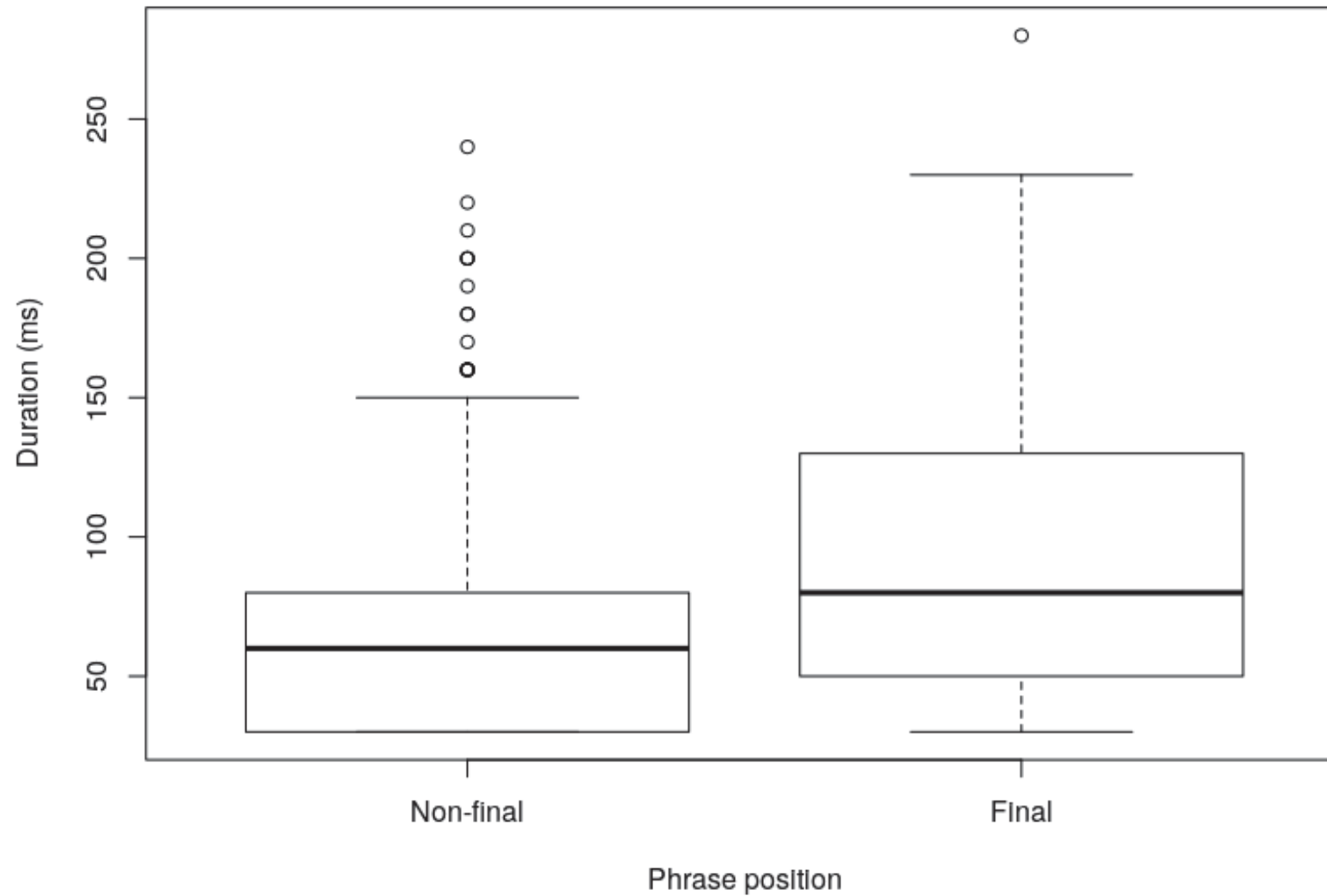
Reminder: the **SVLR** is the pattern of vowel duration in Scottish English according to which some vowels are:

- **Long** before voiced fricatives, /r/ and morpheme boundaries (*breathe, beer, bee, bees*)
- **Short** elsewhere (*bead, beef, bean*).

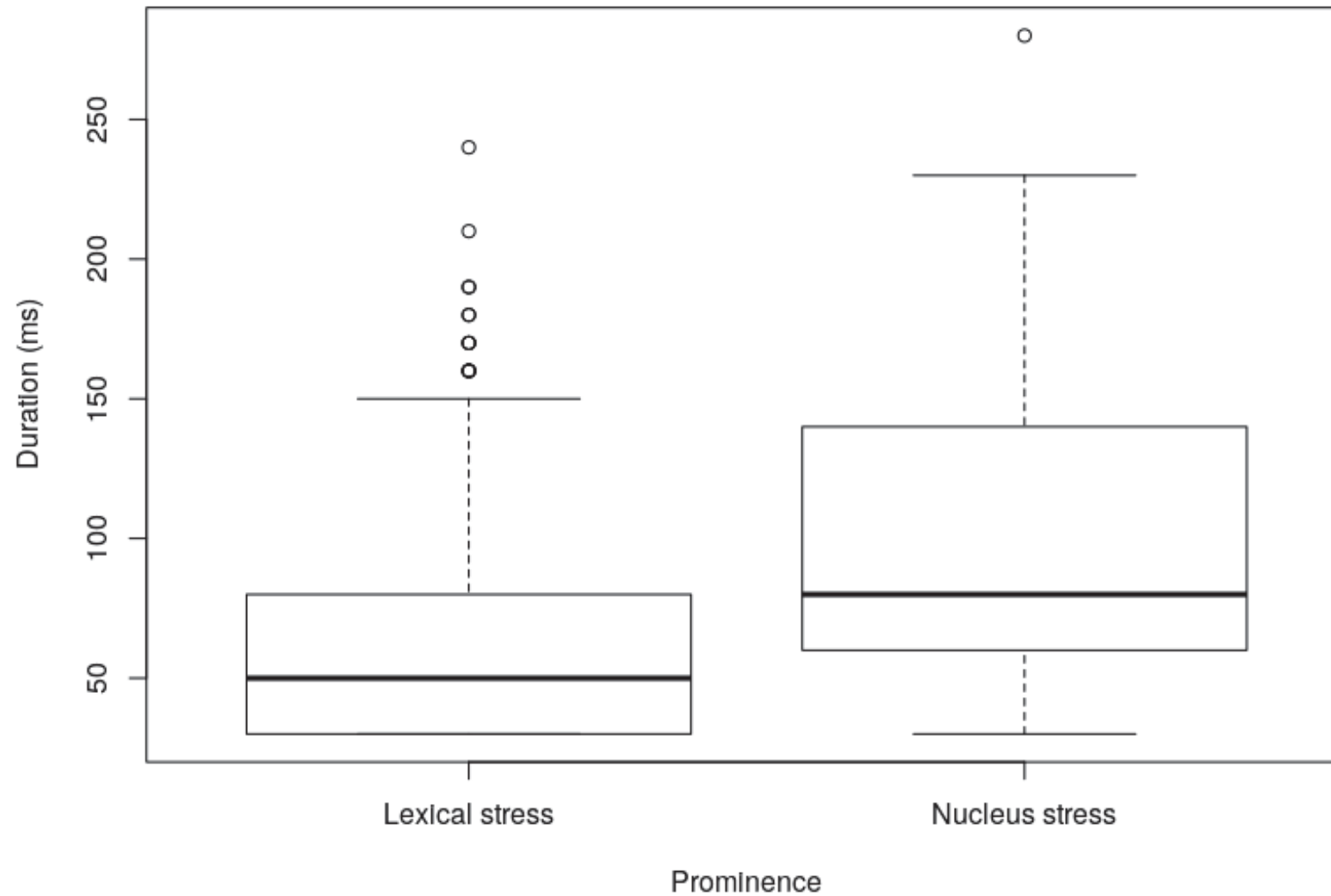
First pilot: presentation

- **Segment durations in ms** were extracted from the corpus during the LaBB-CAT search
- All tokens' prosodic properties (**phrase position, prominence**) were labelled in Praat
- Expected length in **SVLR** pattern was coded based on following morpho-phonemic context (short, phonologically long, morphologically long)
- Other linguistic factors were included: **word type, number of syllables in the word, number of segments in the syllable**

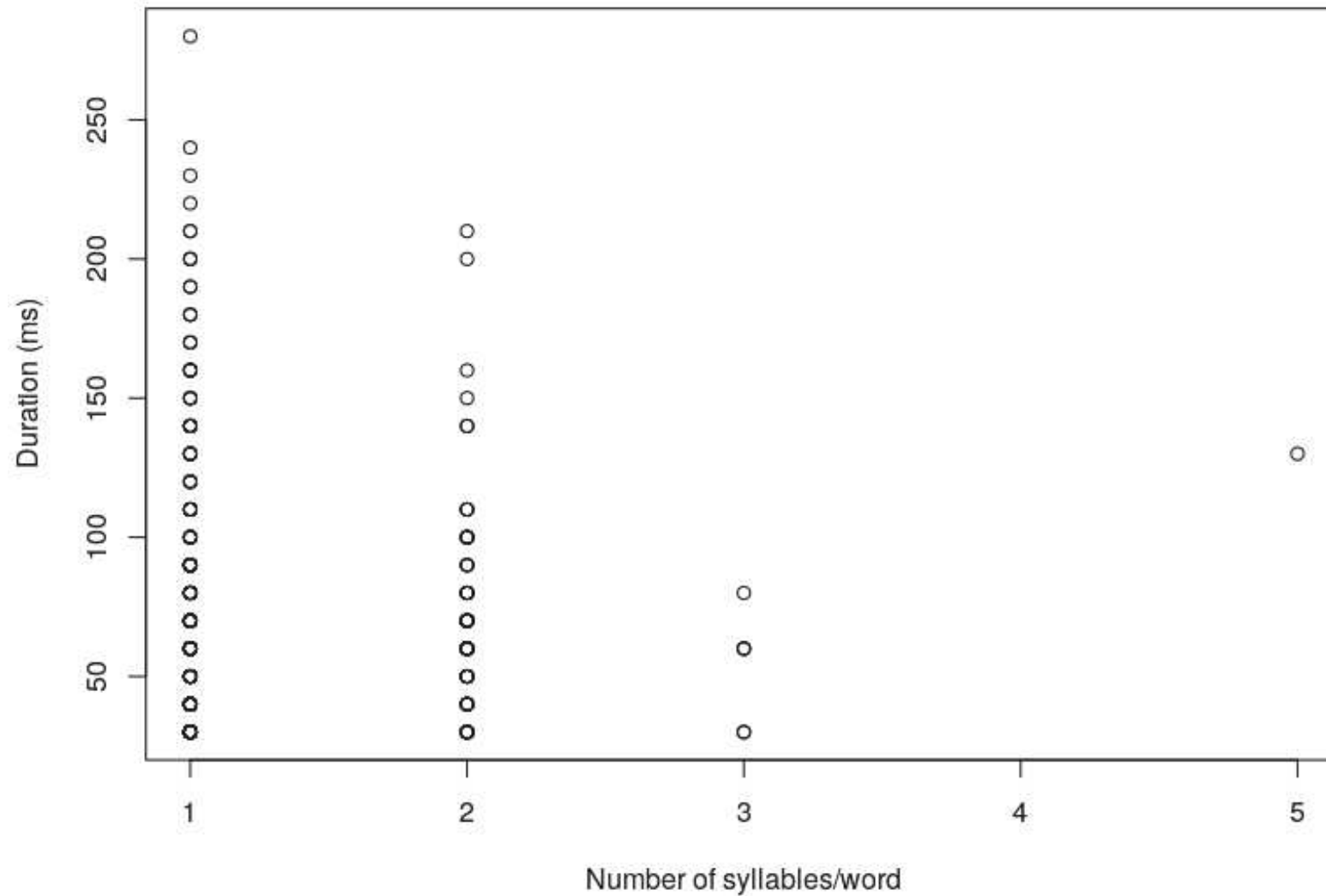
First pilot: partial results



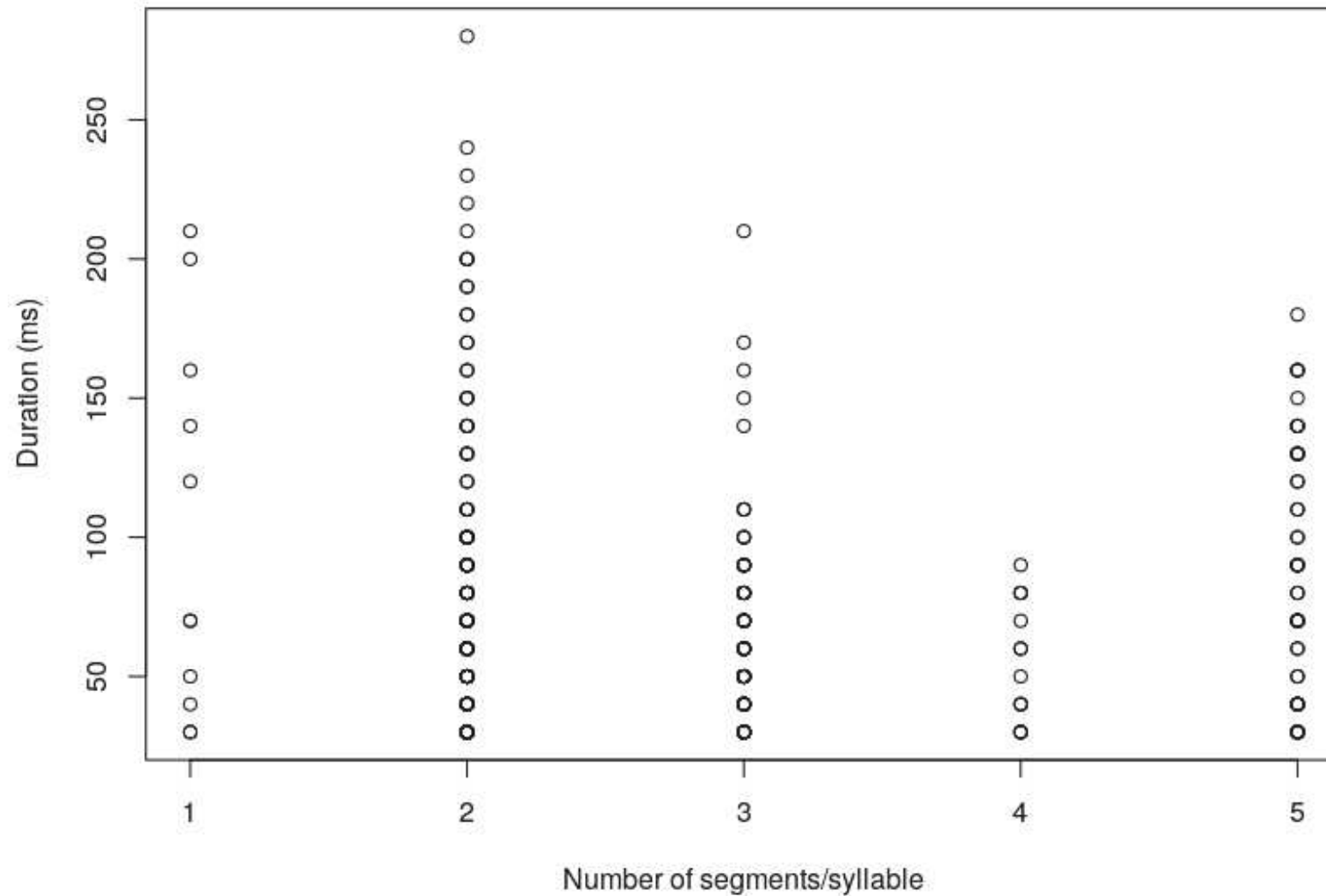
First pilot: partial results



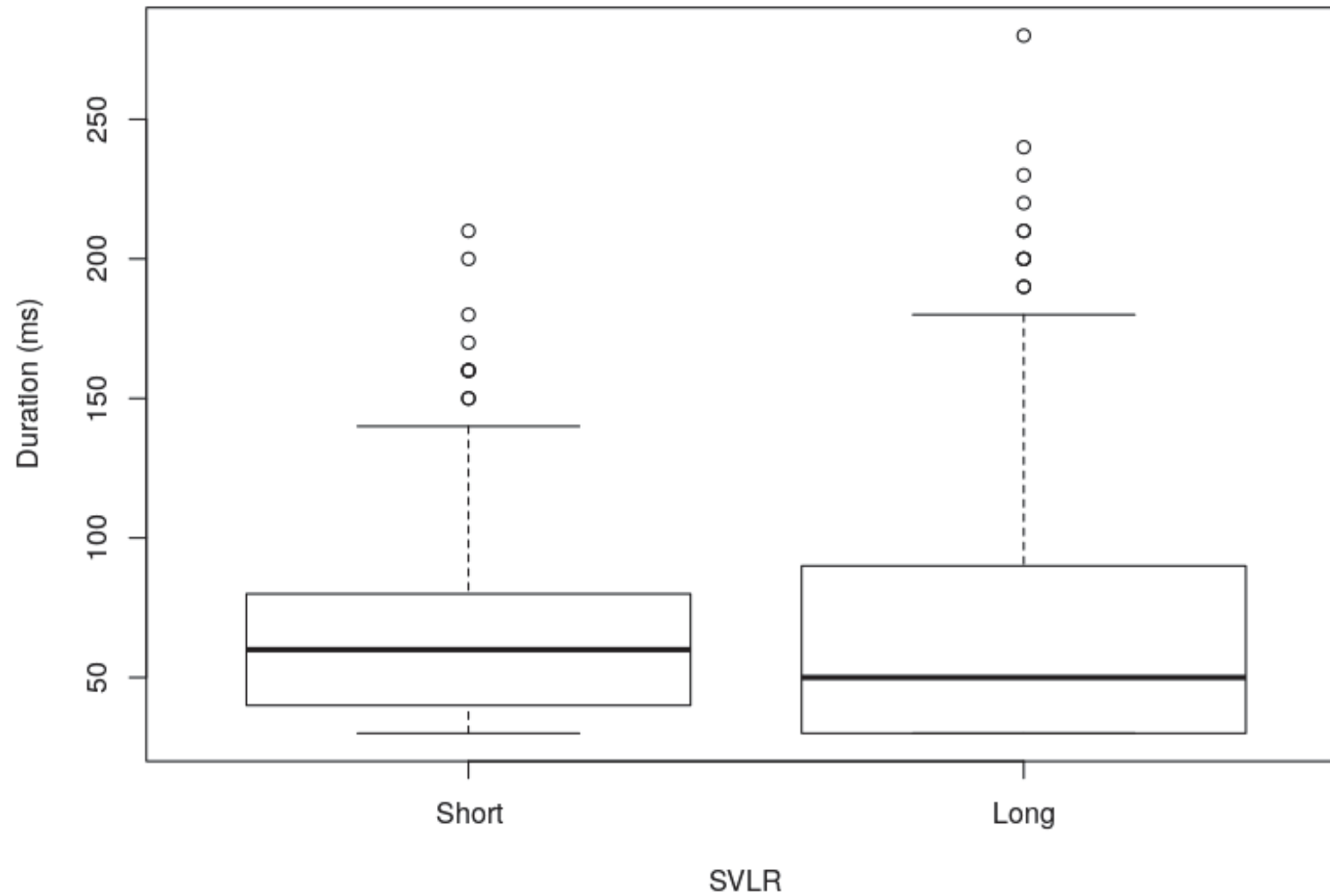
First pilot: partial results



First pilot: partial results



First pilot: partial results



First pilot: partial results

Statistical analysis conducted in R (v 3.5).

- Using linear mixed effects modelling: *lmer*, *step* and *ls_means* with *lme4/lmerTest* packages.
- Random factors: speaker, word.
- Fixed factors: SVLR, phrase position, prominence ; plus all possible interactions.
- Covariates: nb of segments/syll, nb of syll/word.

First pilot: partial results

- Significant factors : **SVLR ***, **phrase position *****, **prominence *****
- Interactions: **svlr:phrase** and **svlr:prominence**

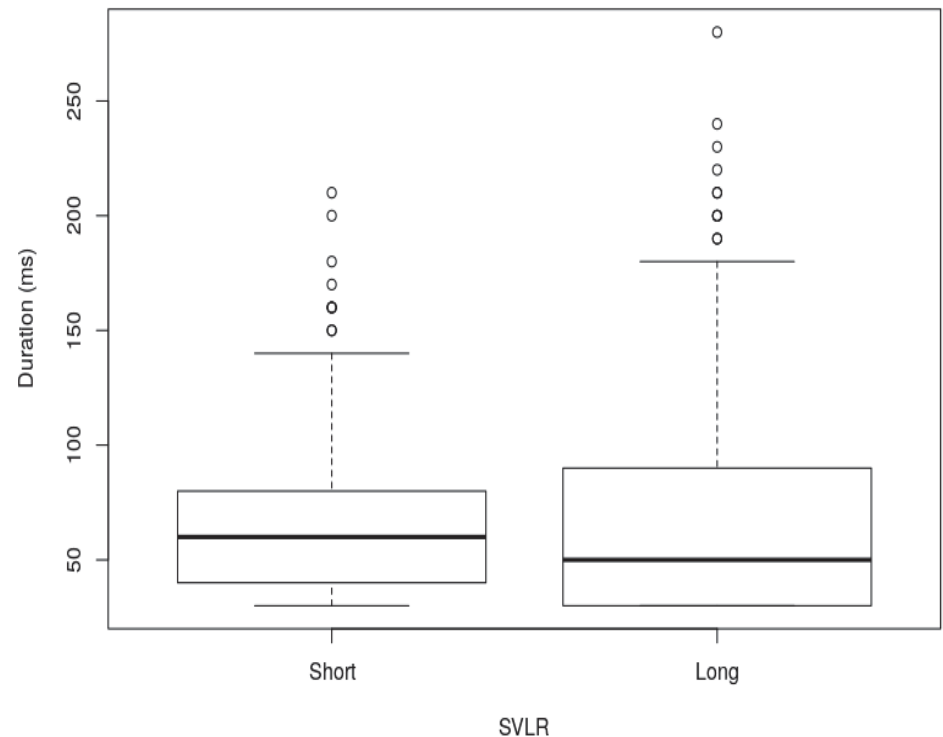
SVLR-short: non-final/final	
SVLR-long: non-final/final	***
Non-final position: short/long	*
Final position: short/long	***

SVLR-short: accent/nucleus	*
SVLR-long: accent/nucleus	***
Accented: short/long	.
Nucleus: short/long	***

- But dynamic variation is still not accounted for in the modelling...

First pilot: issues raised

- Prosody will have to be coded by hand
- How to include time/turn taking into data visualisation and modelling?
- Concerns about the segmentation: **SVLR** *short/long* difference is not very significant ($p = 0.0276$)

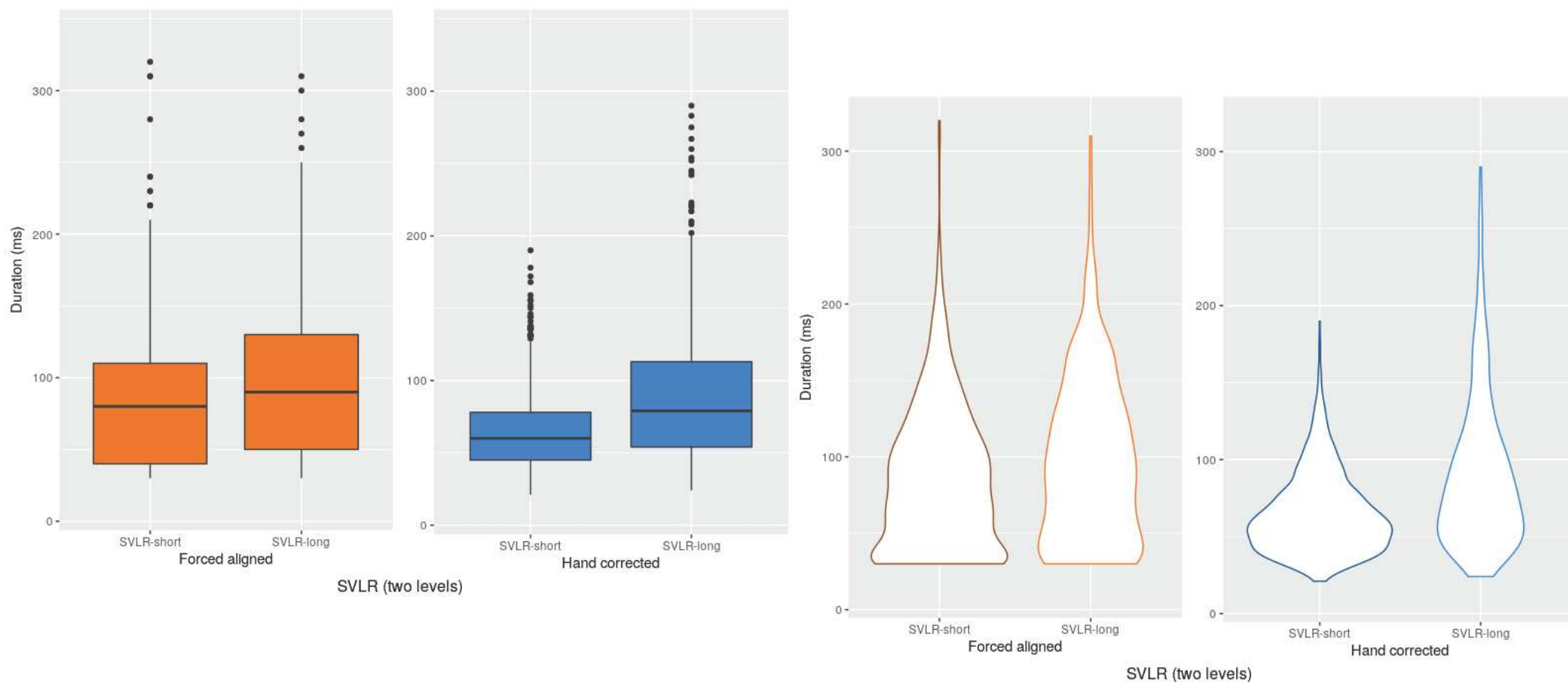


The issue of segmentation: automated vs manual

- How can I know whether I can rely on this corpus' forced-aligned temporal measurements?
- Digging out older data from the same corpus about SVLR (Chevalier 2016): 12 speakers (all female, 4 age groups), two segments (/i/ and /u/), N = 1441
- Similar labelling for phonetic/linguistic factors
- But the segmentation had been hand-corrected for all tokens

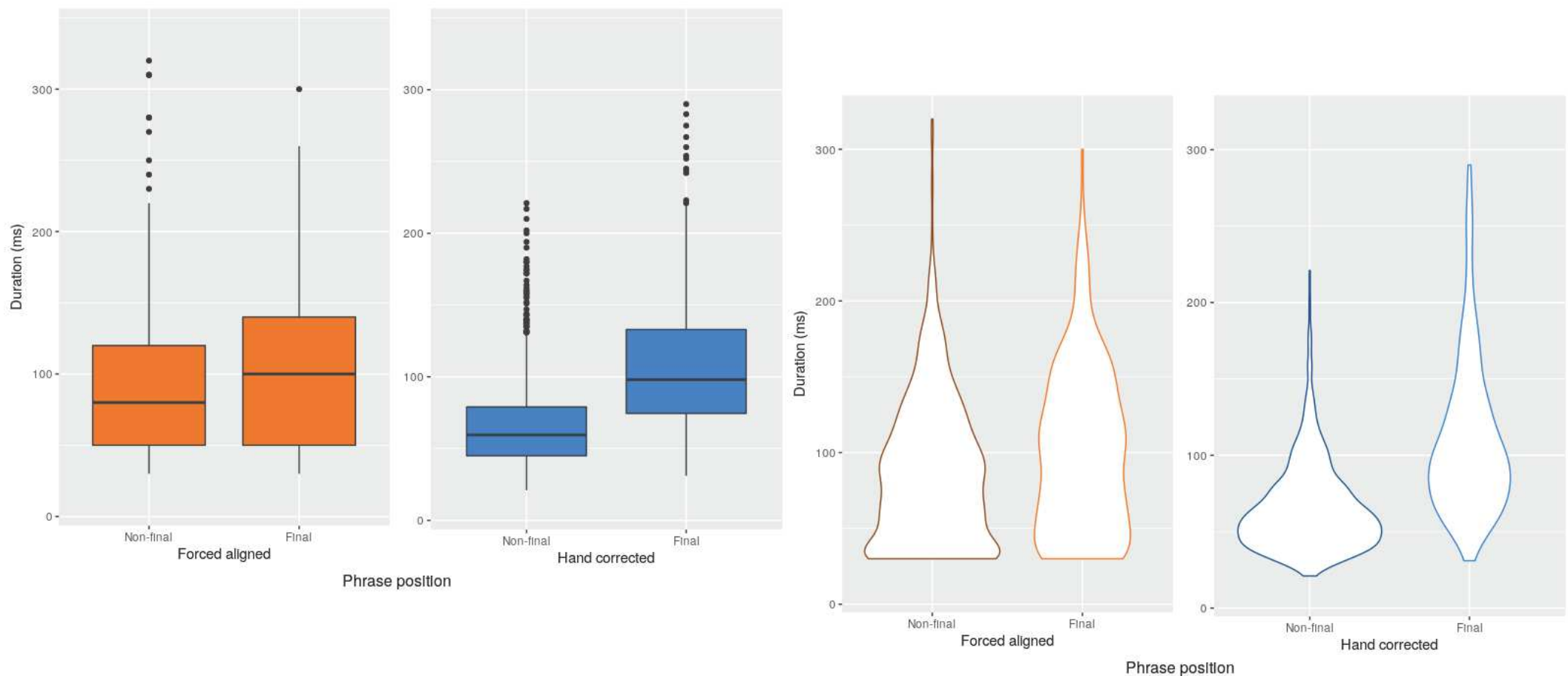
Auto/manual segmentation: Masters' work data

What was the **SVLR** *short/long* distinction like in my previous work on the same corpus?



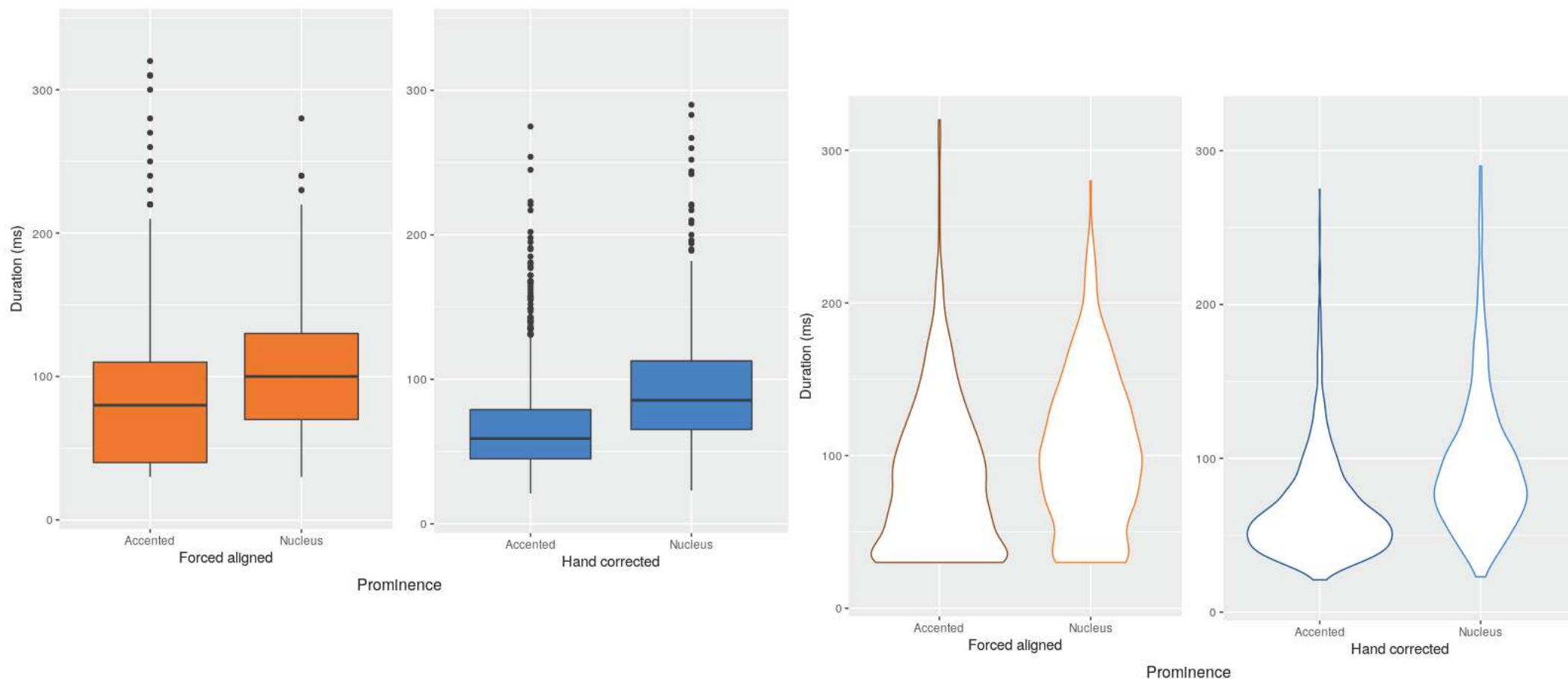
Auto/manual segmentation: Masters' work data

How about **phrase position** (i.e. *non-final/final* opposition)?



Auto/manual segmentation: Masters' work data

How about **prominence** (i.e. *accented/nucleus* opposition)?



Auto/manual segmentation: Masters' work data

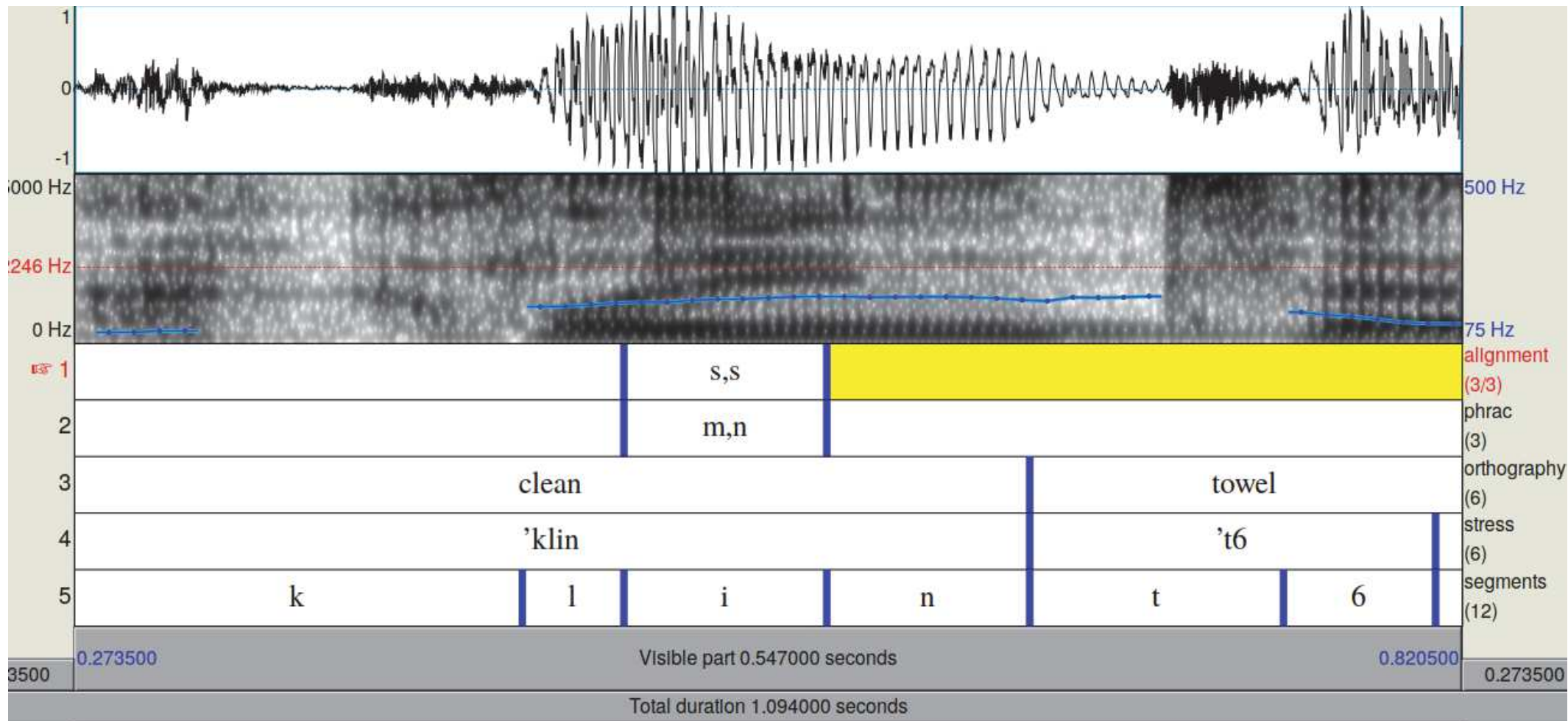
In terms of statistical analysis:

- Information about preceding and following segment have been added but several low counts situations = impossible to use these factors in a model
- Models have been run with SVLR, phrase position, prominence, etc = different outputs
- ANOVA: the model with hand-corrected durations seems better (BIC of 13706 vs 15141)
- I will have to hand-correct some (if not all!) tokens ... But which ones?

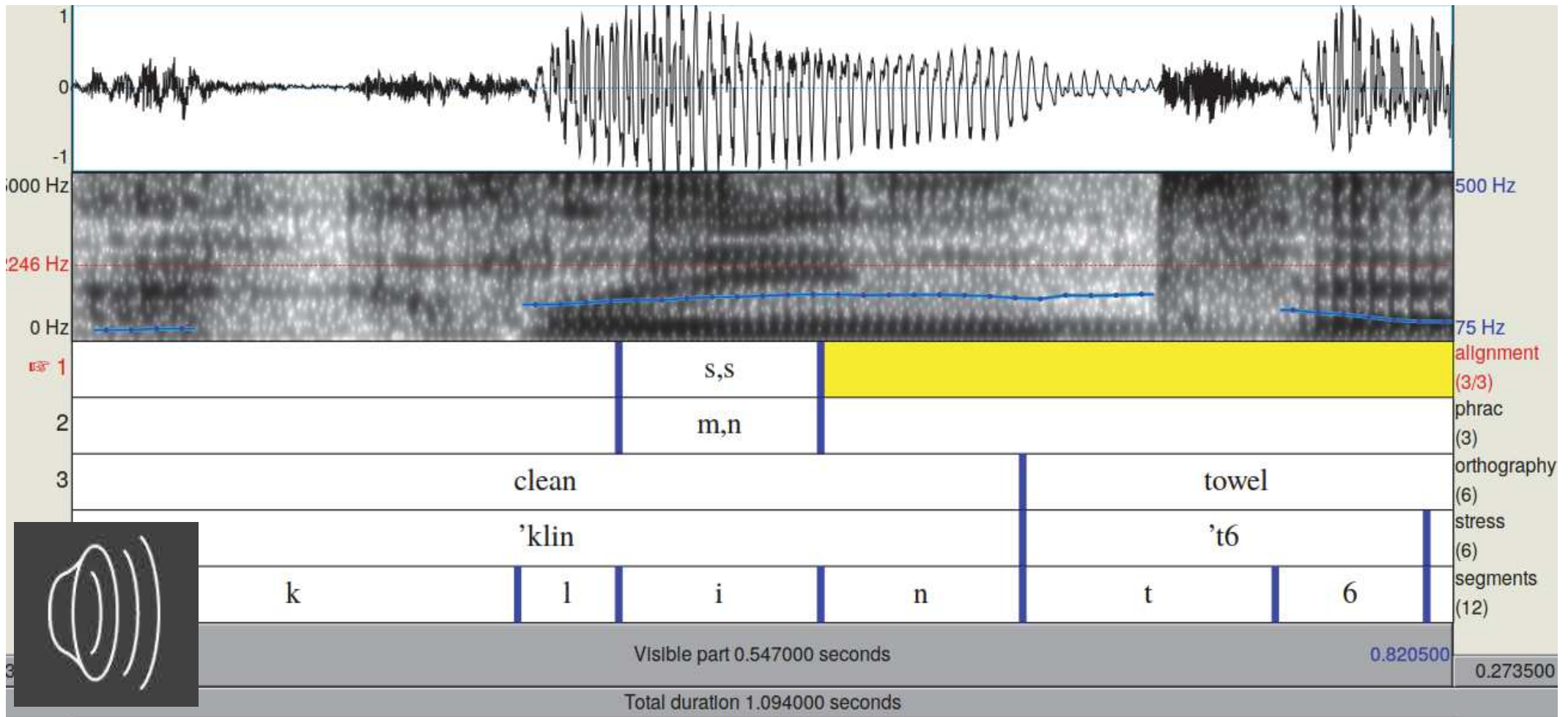
Auto/manual segmentation: first pilot sample

- Now trying to take that experiment further with my firstpilot sample, which has not been hand-corrected at all yet.
- Accuracy of forced-aligned start and end interval boundaries was labelled in Praat.
- 6 levels of accuracy: Extremely early, slightly early, Acceptable, slightly late, extremely late, nowhere close.

Auto/manual segmentation: first pilot sample



Auto/manual segmentation: first pilot sample



Auto/manual segmentation: first pilot sample

Position of segment's interval start/end boundaries

Interval boundary ...	Extremely early	Slightly early	Acceptable	Slightly late	Extremely late	Nowhere close
<i>Start</i> N =	109	126	144	4	11	19
<i>Start</i> (as %)	26.4 %	30.5 %	34.9 %	1.0 %	2.7 %	4.6 %
<i>End</i> N =	56	116	207	5	10	19
<i>End</i> (as %)	13.6 %	28.1 %	50.1 %	1.2 %	2.4 %	4.6 %

Auto/manual segmentation: first pilot sample

Position of segment's interval start/end boundaries

Interval boundary ...	Early	Acceptable	Late	Nowhere close
<i>Start</i> N =	109	274	11	19
<i>Start</i> (as %)	26.4 %	66.3 %	2.7 %	4.6 %
<i>End</i> N =	56	328	10	19
<i>End</i> (as %)	13.6 %	79.4 %	2.4 %	4.6 %

Auto/manual segmentation: first pilot sample

Start = Prev Seg	Early	Ok	Late
Voiceless obstruent	22.8 %	73.7 %	1.8 %
Voiceless fricative	8.0 %	85.8 %	3.6 %
Voiced obstruent	60.0 %	31.1 %	/
Nasal	53.5 %	34.9 %	2.3 %
Lateral	33.3 %	66.7 %	/
Voiced fricative	15.8 %	65.8 %	5.3 %
Morph. boundary	29.4 %	58.8 %	11.8 %
r	11.9 %	88.1 %	/
glide	66.7 %	19.0 %	4.8 %

Start interval bound./preceding segment

End = Next seg	Early	Ok	Late
Voiceless obstruent	7.9 %	90.1 %	2.0 %
Voiceless fricative	14.3 %	85.7 %	/
Voiced obstruent	36.4 %	63.6 %	/
Nasal	14.5 %	75.4 %	2.9 %
Lateral	10.0 %	90.0 %	/
Voiced fricative	25.0 %	68.8 %	6.3 %
Morph. boundary	14.1 %	76.4 %	2.5 %

End interval bound./next segment

Auto/manual segmentation: first pilot sample

Start and end interval boundaries: **speaker, word type**

<i>START</i>	Early	Ok	Late	Bad
90M m01	28.4 %	66.9 %	2.1 %	2.6 %
90M m02	16.7 %	63.9 %	5.6 %	13.9 %

<i>END</i>	Early	Ok	Late	Bad
90M m01	14.1 %	81.5 %	7.8 %	2.6 %
90M m02	11.1 %	69.4 %	5.6 %	13.9 %

<i>START</i>	Early	Ok	Late	Bad
Funct.	32.3 %	56.2 %	3.5 %	8.0 %
Cont.	20.8 %	75.9 %	1.9 %	1.4 %

<i>END</i>	Early	Ok	Late	Bad
Funct.	15.9 %	72.6 %	3.0 %	8.5 %
Cont.	11.3 %	85.8 %	1.9 %	0.9 %

Auto/manual segmentation: first pilot sample

Start and end interval boundaries: **prosodic position**

<i>START</i>	Early	Ok	Late	Bad
Initial	33.3 %	57.8 %	6.7 %	2.2 %
Median	14.6 %	68.0 %	2.0 %	5.4 %
Final	29.6 %	64.8 %	2.8 %	2.8 %

<i>END</i>	Early	Ok	Late	Bad
Initial	22.2 %	68.9 %	6.7 %	2.2 %
Median	12.5 %	81.1 %	1.4 %	5.1 %
Final	12.7 %	78.9 %	4.2 %	4.2 %

<i>START</i>	Early	Ok	Late	Bad
Stress	34.1 %	36.6 %	4.9 %	24.4 %
Accented	23.8 %	71.0 %	2.3 %	3.0 %
Nucl.	33.3 %	63.8 %	2.9 %	/

<i>END</i>	Early	Ok	Late	Bad
Stress	17.1 %	56.1 %	4.9 %	22.0 %
Accented	13.9 %	81.5 %	1.7 %	3.0 %
Nucl.	10.1 %	84.1 %	4.4 %	1.5 %

Next steps

Food for thought:

- General advice about accuracy of automated segmentation in SotC/HTK-based/any forced-aligned corpus when studying temporal features?
- Any ways of identifying the contexts in which automated segmentation is fine? (e.g. Modelling with a binary value auto/hand as dependent variable and including preceding/following segment)
- Visualising the progression of the conversation?

THANK YOU

(and enjoy your meal)