Seminar session 2

Gemination and Degemination in English Affixation

Lexical Strata, Variability, and Phonetic Evidence

Collaborators
Sonia Ben Hedia (U Düsseldorf), Gero Kunter (U Düsseldorf)
Funding
Strategischer Forschungsförderfonds Heinrich-Heine-Universität Düsseldorf
Deutsche Forschungsgemeinschaft
- Grant PL151/8-1 'Morpho-phonetic Variation in English'
- Grant PL151/7-1 'FOR 2737 Spoken Morphology: Central Project'

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

DFG Deutsche Forschungsgemeinschaft

Labex EFL SORBONNE PARIS CITÉ

Morpho-phonology as we know it

- Morpho-phonological alternations are categorical but may have lexical exceptions.

- The formal level of representation of morphemes is phonological in nature.

- Post-lexical phonology and phonetics have no access to lexical information.

Problems

- Morpho-phonological alternations are more variable than previously assumed, and governed by unexpected factors
  - Stress shift (Bauer, Lieber & Plag 2013 on *-able*)
  - Stress preservation (Collie 2008, relative frequency as a proxy for morphological segmentability)

- Subphonemic detail may reflect morphological information
  - Free vs. bound stems (Kemps et al. 2005, Blazej & Cohen-Goldberg 2015)
  - Paradigmatic enhancement (Cohen 2014, Kuperman et al. 2005)

- Serious implications for theories of morpho-phonology (Plag 2014)
  - Exception vs. rule
  - Lexical vs. post-lexical phonology (in linguistic theory, and in speech production models)

Problems

- Morpho-phonological alternations are more variable than previously assumed, and governed by unexpected factors
  - Stress shift (Bauer, Lieber & Plag 2013 on *-able*)
  - Stress preservation (Collie 2008, relative frequency as a proxy for morphological segmentability)

- Subphonemic detail may reflect morphological information
  - Free vs. bound stems (Kemps et al. 2005, Blazej & Cohen-Goldberg 2015)
  - Different S morphemes (Plag, Homann & Kunter 2015)

- Serious implications for theories of morpho-phonology (Plag 2014)
  - Exception vs. rule
  - Lexical vs. post-lexical phonology (in linguistic theory, and in speech production models)

Consonant gemination

- Length contrast of consonants
- Phonetic correlate:
  Gemminate has longer duration than singleton
  (relative or absolute duration)

Italian: *cane* ‘dog’ *canne* ‘canes’

Arabic: *kasara* ‘he broke’ *kassara* (‘he smashed’)

English: *some more*, *room mate*, *unknown*
only across morphemic boundaries
‘fake gemination‘, ‘morphological gemination‘
‘gemination‘

Views on gemination in English

- “Double consonants are frequently found in English, especially at word junction: *wholly* (as said by many), *unknown*, *book-case*, *this Sunday*.” (Abercrombie1967:82)

- “geminate consonants occur in English only at morpheme boundaries: *nighttime*, *bookcase*, *solely*, *non-null*.” (Trask 1996:154)

- “In a more formal, careful speech style, some native speakers may geminate some words, as Trask (*op. cit*.) notes. Some of these for some native speakers might, in fact, be spelling pronunciations. Thus, a word such as *unknown* may actually be pronounced by some with a geminated [nn] due to the pronunciation of its orthographic representation. A geminated [nn] in *unknown*, however, sounds awkward in my own speech, but there is always the possibility of a pragmatically based, purposeful gemination, i.e., for special effect.” (Kaye 2005)

- Dictionaries vary a great deal w.r.t. the transcription of pertinent words with one or two phonetic symbols (see Kaye 2005 for illustration)

Views on gemination in English

Table 9. *Geminate consonants over morphological boundaries*

Geminate	In a compound	In affixation
bb	deadbeat	
dd	god-daughter	mid-day
kk	sack cloth	
ll	full-length	goalless, cruelly
mm	film-maker	alignment, in-mate, embalmment
nn	grand-niece	unknown, drunkenness
pp	lamp-post	step-parent
rr[20]	ear-ring	fore-runner
ss	cross-section	dissatisfaction, ex-service
tt	test-tube	

(from Bauer 2015:30)

Gemination with *-ly*

Adverbial –*ly*

gemination: stalely, vilely

no gemination: fully really

variable gemination: dully, wholly

(http://dict.leo.org, Wells 2013)

Statements about *un-* and *in-*

Mohanan (1986:18):

degemination: *innavigable* and *innumerable* (level 1)
gemination: *unknown*, *unnatural*, *suddenness*, *fineness*, *soulless*, *guileless*.
(level 2)

Gimson's Pronunciation of English (2014):

"In general such prefixes result in a doubled consonant when the prefix-final and the stem-initial consonants are identical, e.g. *unnecessary* is pronounced with a double length [n:]. (This rule does not apply to *in-* and its variants, so for example *illogical* is pronounced with only a single /l/)." (p. 248)

Cohen-Goldberg (2013: 1055f):

"Similarly in English, although geminates are banned from monomorphemic words (*spaghe[tt]i) and words containing less productive affixes (e.g.in-: i[n]umerable), they are allowed in words containing more productive affixes and compounds (e.g. un-: u[nn]ecessary; boo[kk]eeper)."

Hypotheses

- a difference between lento speech (formal) and allegro speech (informal)
  (Kaye 2005)
- *un-* geminates
  (e.g. Cruttenden & Gimson 2014, Cohen-Goldberg 2014, Kiparsky 1982, Mohanan 1986)
- *in-* degeminates
  (e.g. Cruttenden & Gimson 2014, Cohen-Goldberg 2014, Kiparsky 1982, Mohanan 1986)
- *-ly*
  - ... is variable (*stalely* vs. *fully*, Bauer 2001, Bauer, Lieber & Plag 2013)
  - ... geminates (Lexical Phonology: level 2 affix)

- General theoretical assumptions by morpho-phonologists
  - Degemination is affix- or stratum-dependent
  - Degemination is a categorical morpho-phonological process with some lexical exceptions

Predictions

	Level 1	**Level 2**
Morphological Process	*in + numerous*	*un + natural* *sole + ly*
Phonological Process	*i*/n/*umerous*	*u*/nn/*atural* *so*/ll/*y*
Phonetic Outcome	*i*[n]*umerous*	*u*[n:]*atural* *so*[l:]*y*

	Degemination	**Gemination**

Empirical evidence?

- Only two studies empirically investigated *in*- and *un*- in English
- Kaye (2005): experiment with very few types, spoken in isolation

in-	*immature – mature*	gemination (but somewhat variable by speaker)
un-	*unaimed – unnamed – named*	gemination

Empirical evidence?

Oh and Redford (2013)

- Experimental study with four types for each prefix

  immovable, immoral, immemorial, immeasured

  unnoticed, unnamed, unnerve, unnail

- Comparison of durations with (assumed) phonological singletons with orthographic doubles (e.g. *immunity*, *immigrational*)

Results

- both *im-* and *un-* geminate, but *un-* more than *im-*

Problems

- Small set of types, presented orthographically, read speech
- A priori classification of stimuli as geminates or non-geminates
- Stimuli only spoken in carrier sentence 'I said ___ again', asking for normal vs. careful speaking style
- Morphological implications not clear

Empirical evidence?

- No empirical study of -*ly*

This study

- What are the facts?

- What determines (de-)gemination at morphological boundaries?

- Three affixes: *un-, in-, -ly*

- Diagnostics: Acoustic duration

- Data: Natural conversational speech

Methodology

- Switchboard Corpus (Godfrey & Holliman 1997)
  - 2430 two sided phone conversations among North American speakers, 240 hours of speech.
  - 3 million word tokens
- Sample of *un- , in-* and *-ly*-affixed words with a double or a single (orthograhic) consonant at the morphological boundary
- ‘affixed‘: The base must be attested outside the derivative with a similar meaning (*unfair*, *implicit – explicit, innocent*)
- For each affix we sampled up to 160 words per category
- only one token of a given type by a single speaker
- For the prefix *in-* only the allomorph /ɪn/ was investigated (with <nn> only *innate*, *innocent* and *innovate* (with some derivatives) are attested in the corpus)

Methodology

- Manual segmentation and acoustic measurements in Praat (Boersma & Weenink 2014)

- Go to PRAAT!

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Gemination may also affect the vowel preceding the geminated segment
(e.g. Ridouane 2010, Miller 1987, Oh and Redford 2011)

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Coarticulation effects

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Coarticulation effects

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Speech rate directly influences the duration of a given segment.

$$\frac{\texttt{Number of segments}}{\texttt{word duration}}$$

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Final lengthening effect

`mid, end, before pause`

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

- `stressed` vs. `unstressed` affix
- `adjacent` / `non-adjacent` to a stress syllable

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

for *–ly:*

syllabic /l/ should be longer

ment[l̩]y ment[əl]y odd[l̩]y

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

More frequent words are produced faster

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

Measure of gradient morphological complexity

The more frequent the derivative vis-à-vis the base, the less complex the word

happyness - happy
discernment - discernment

government - govern
insane - sane

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

un-		
in-	NEGATIVE	*immature*
in-	LOCATIVE	*immigrant*
-ly		

Methodology

- Statistical Analysis: Multiple regression with **duration** as dependent variable and **number of consonants** (single vs. double) as crucial predictor

- Coding of pertinent covariates:
  - Preceding Segment Duration
  - Preceding Segment
  - Following segment
  - Speech Rate
  - Position in utterance
  - Prosodic Structure
  - Syllabicity
  - Word Form Frequency
  - Relative Frequency
  - Affix
  - Semantic Transparency

transparent:

Affix + Base = Derivative
im + possible = *impossible*
NEG + ‘possible‘ = ‘not possible‘

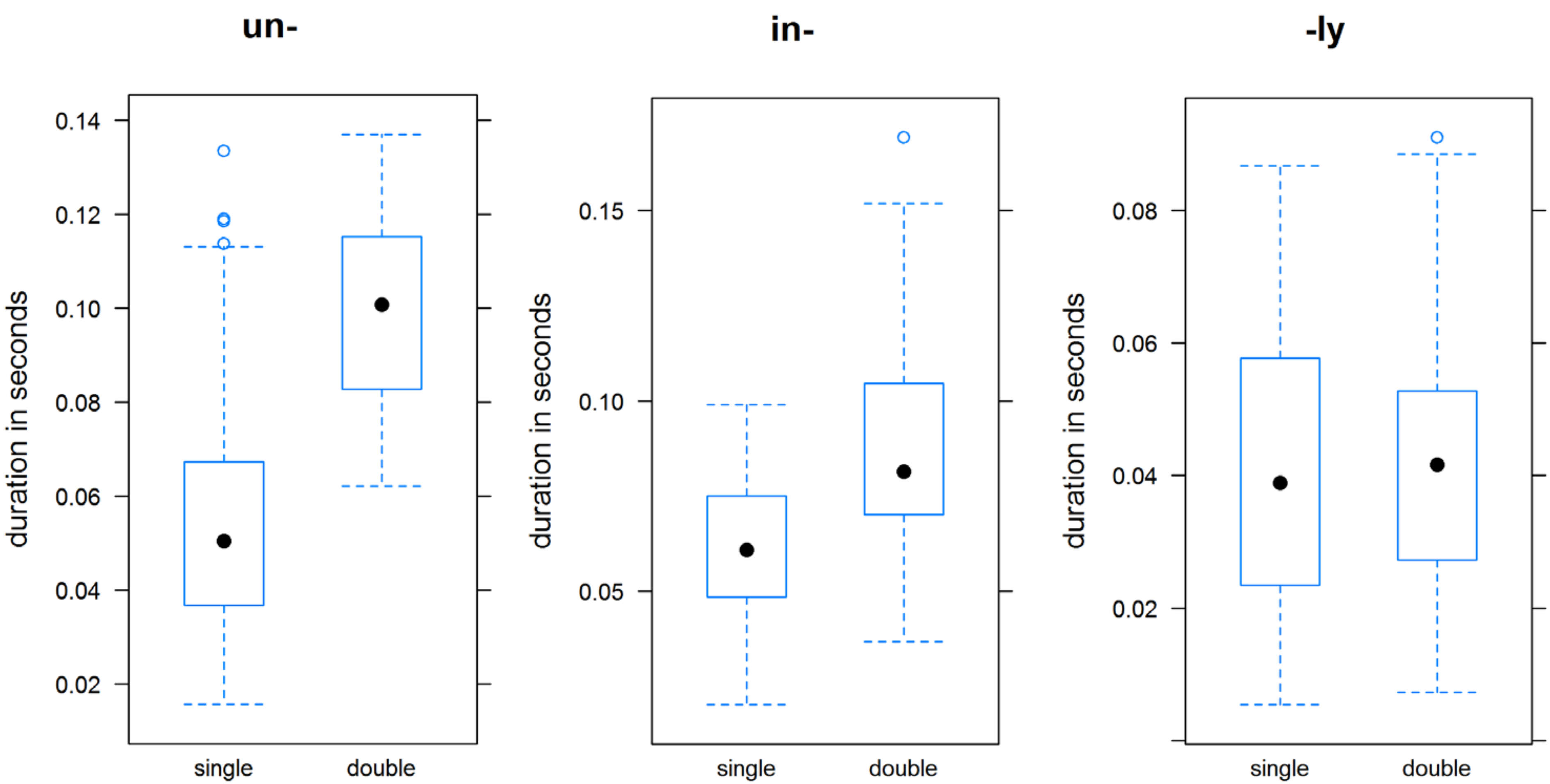
opaque:

im + mediately ≠ ‘at once‘

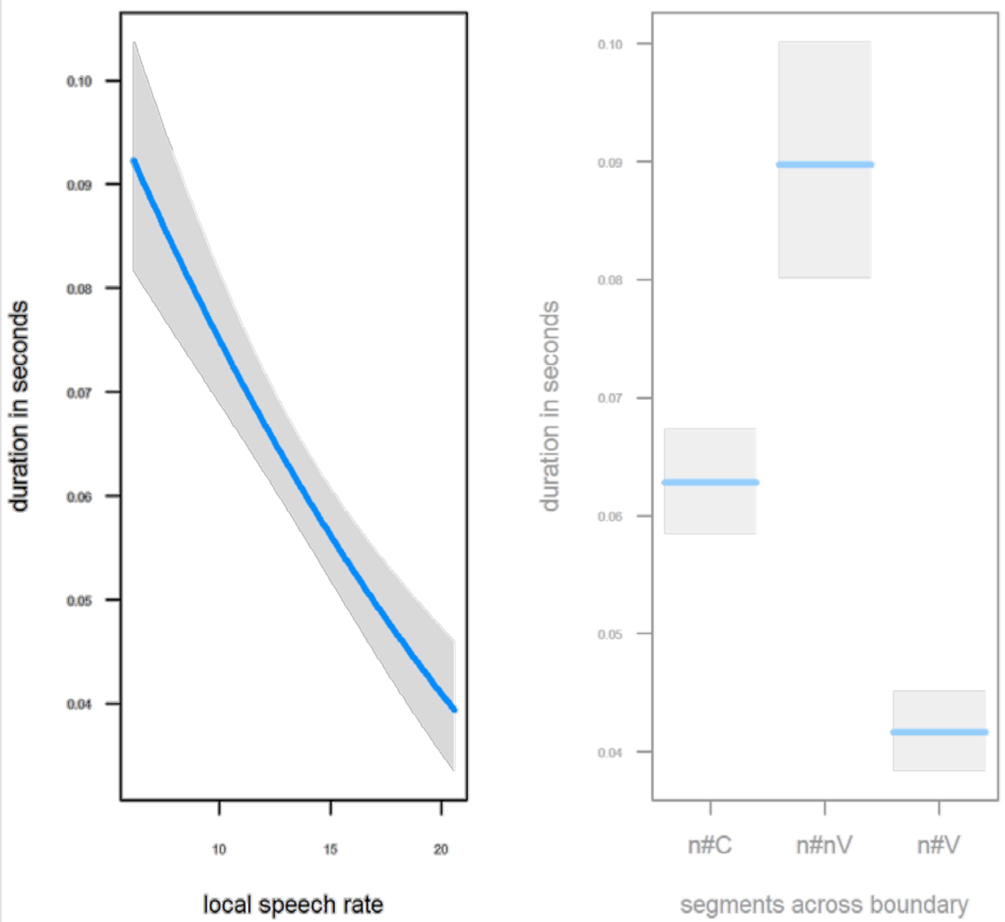
Overview of the data

	Double Consonant	**Single Consonant**	***Total per affix***
un-	22	136	*158*
in-	89	67	*156*
-ly	81	75	*156*

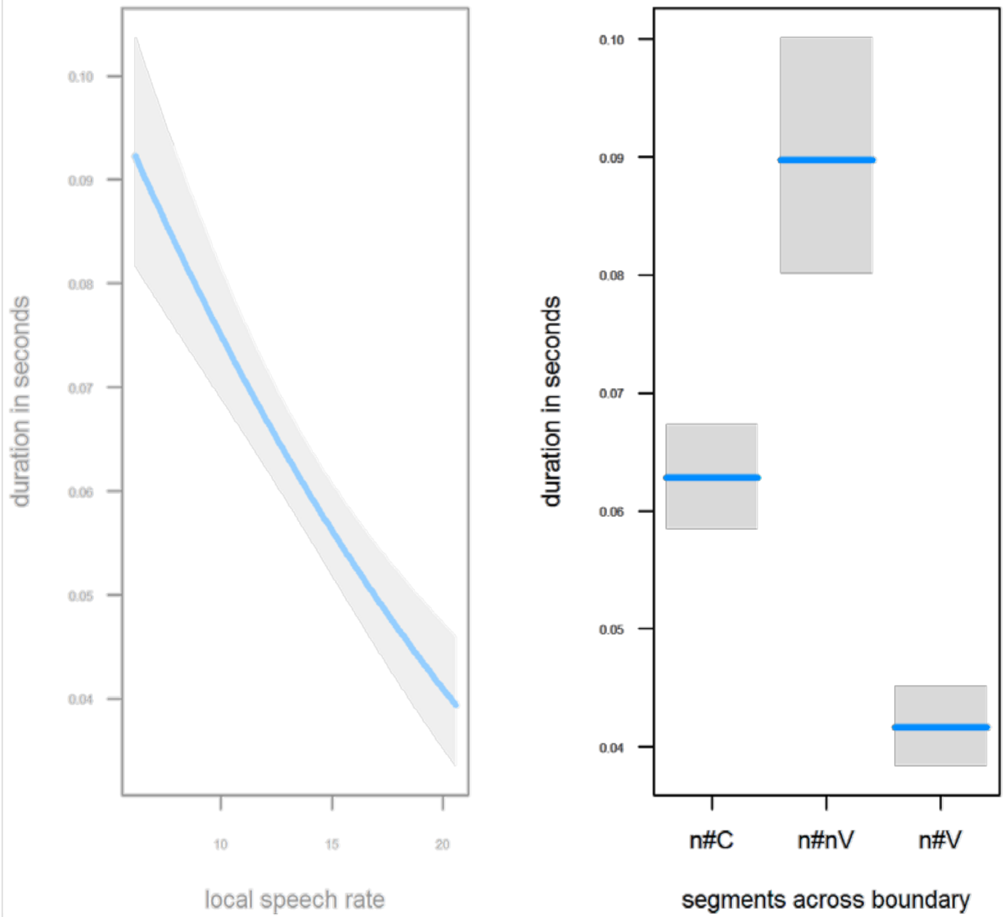
Results: Overview



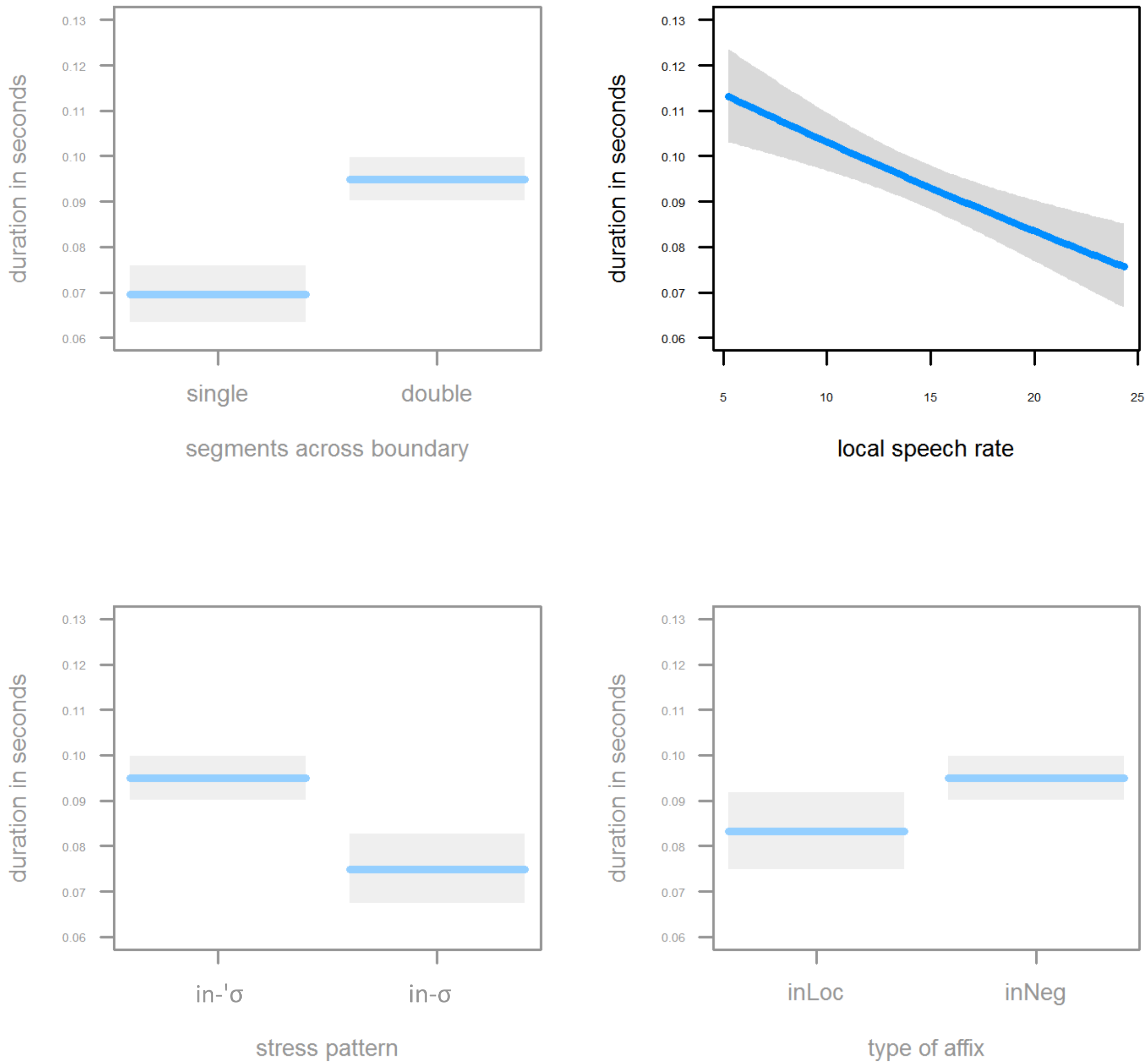
Results 1: *un*- geminates



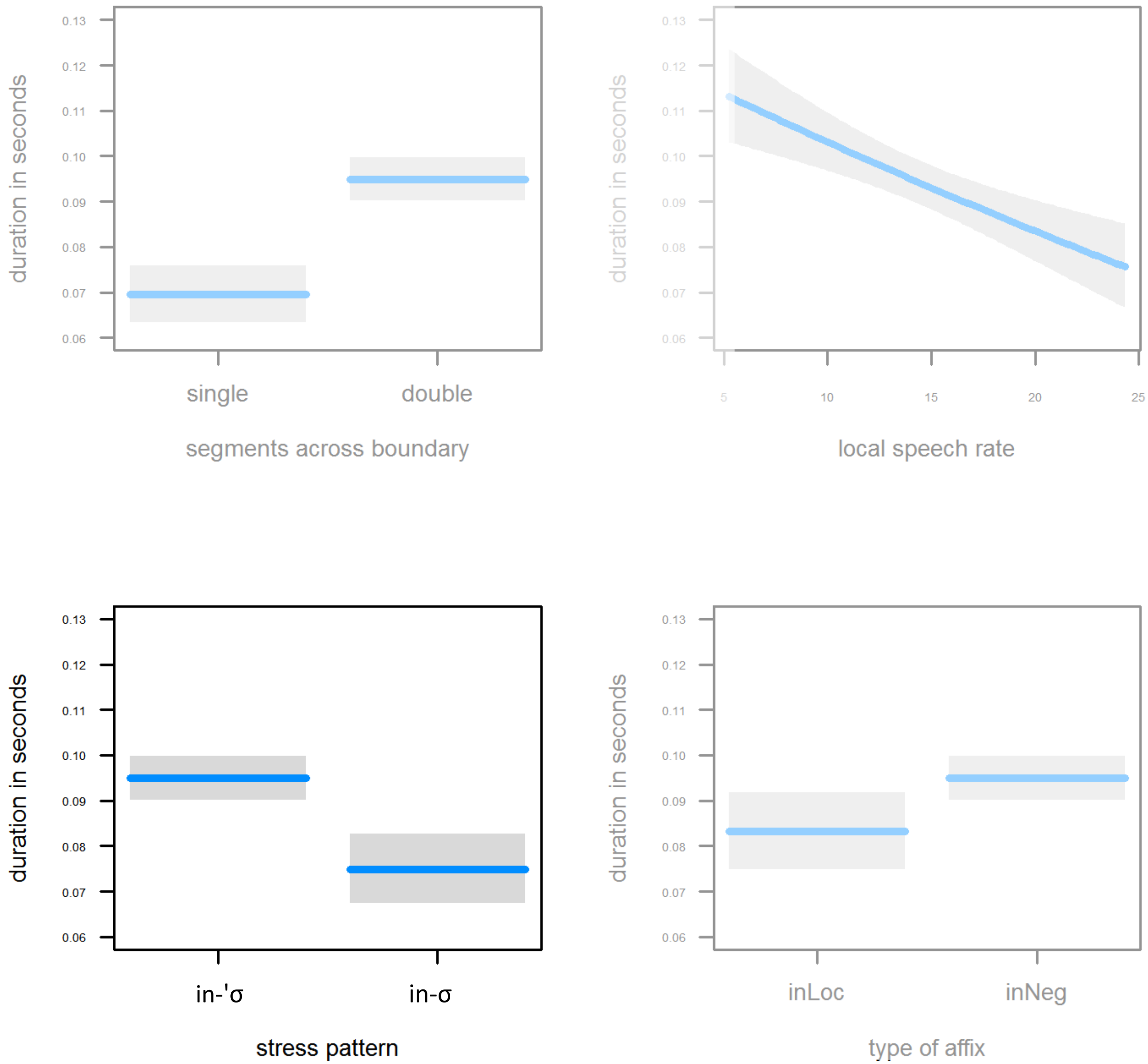
Results 1: *un*- geminates



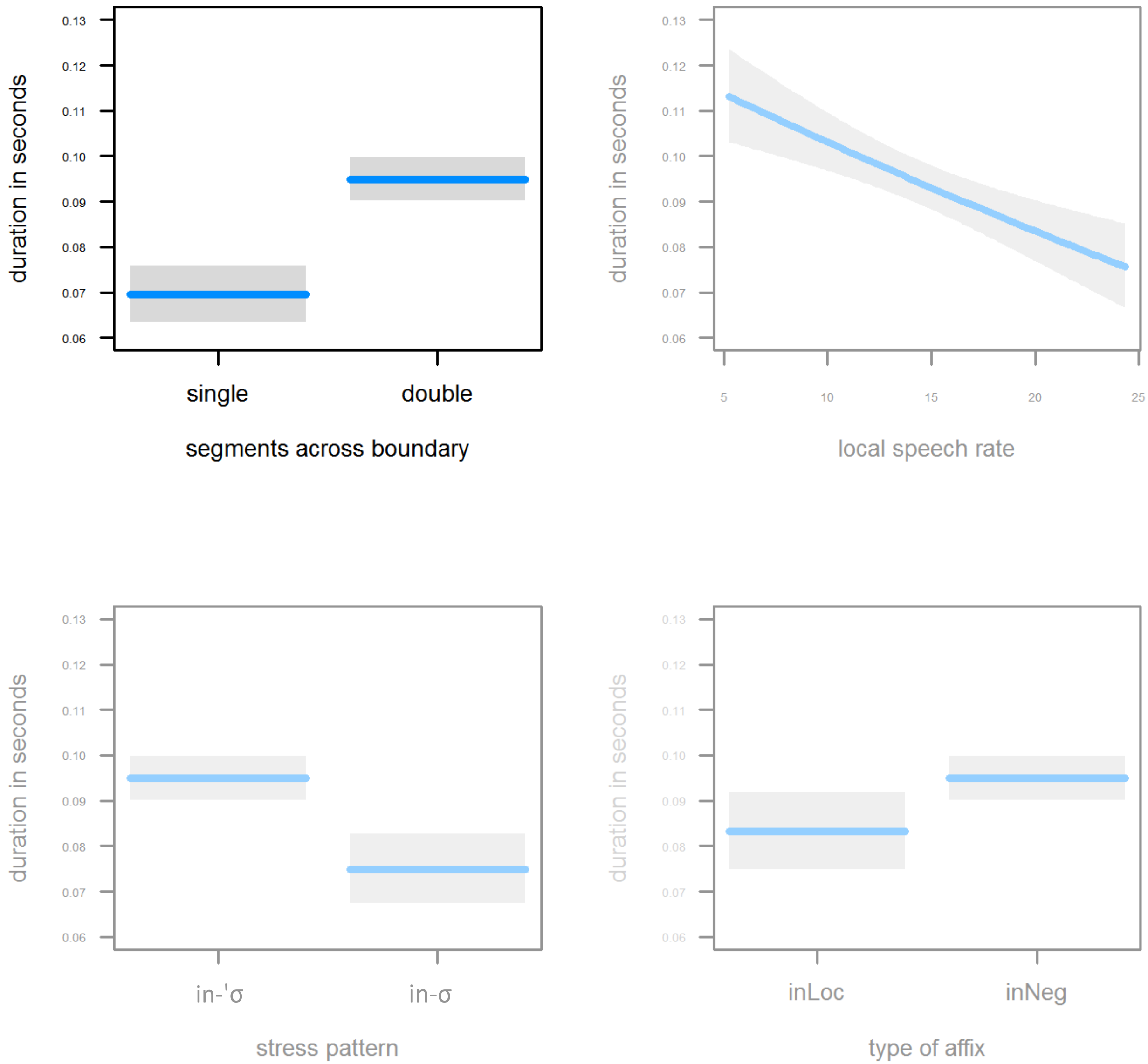
Results 2: *in-* geminates



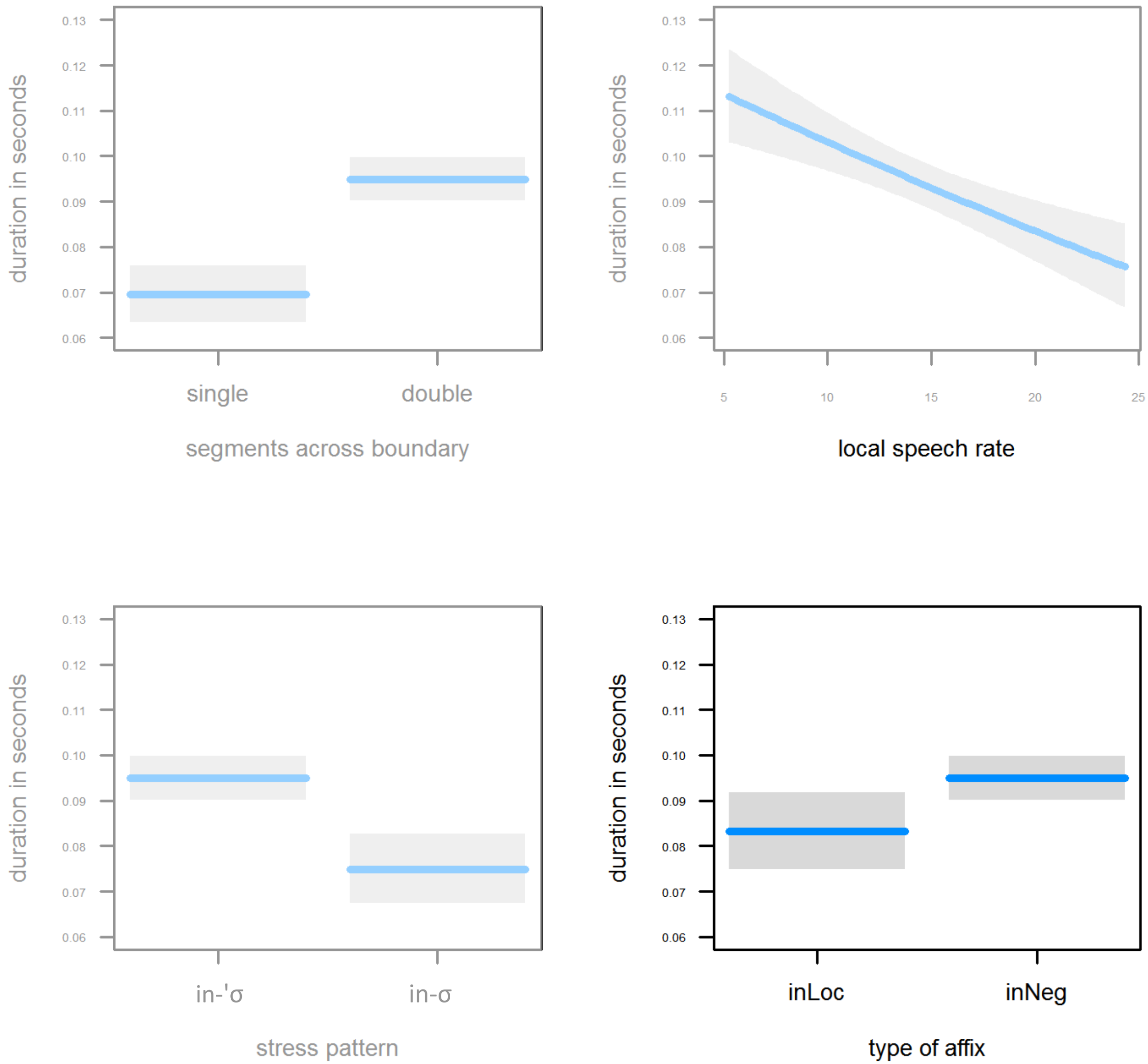
Results 2: *in-* geminates



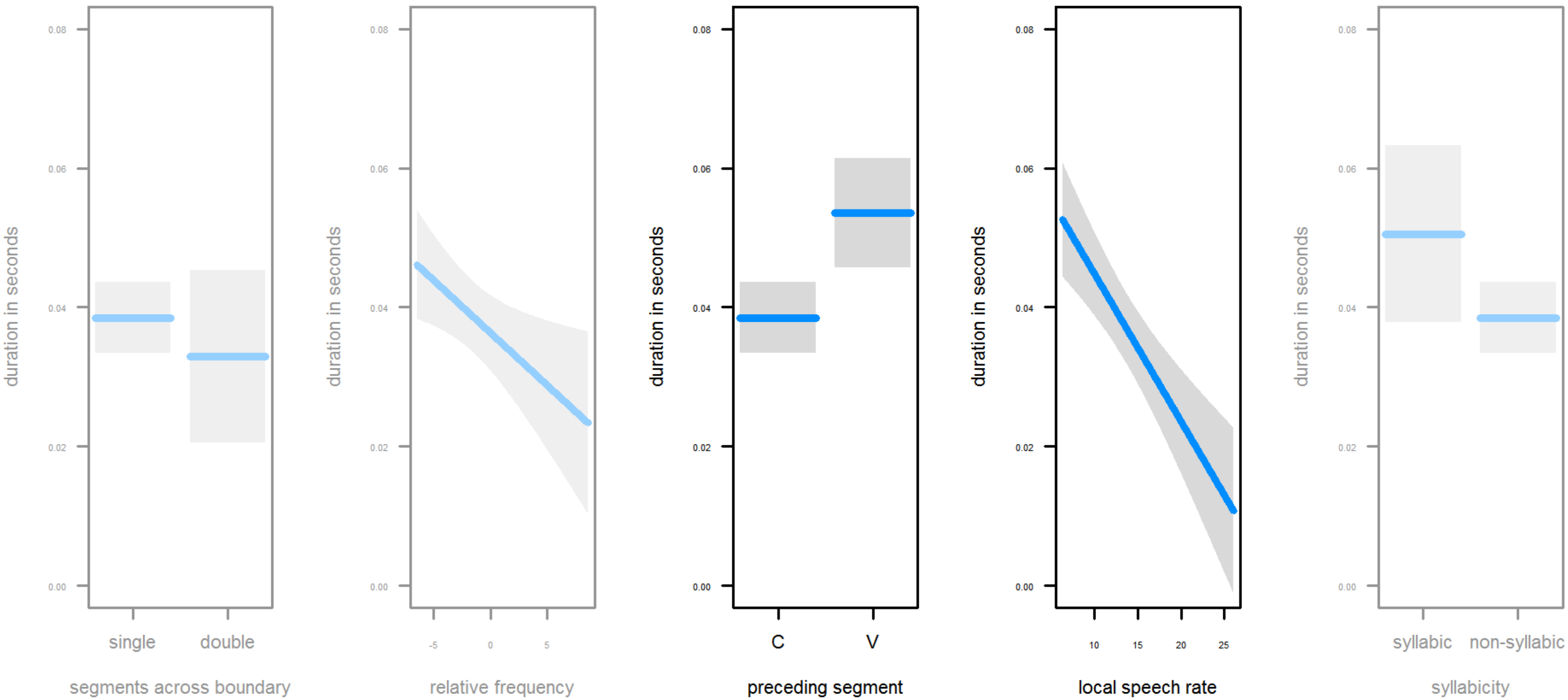
Results 2: *in-* geminates



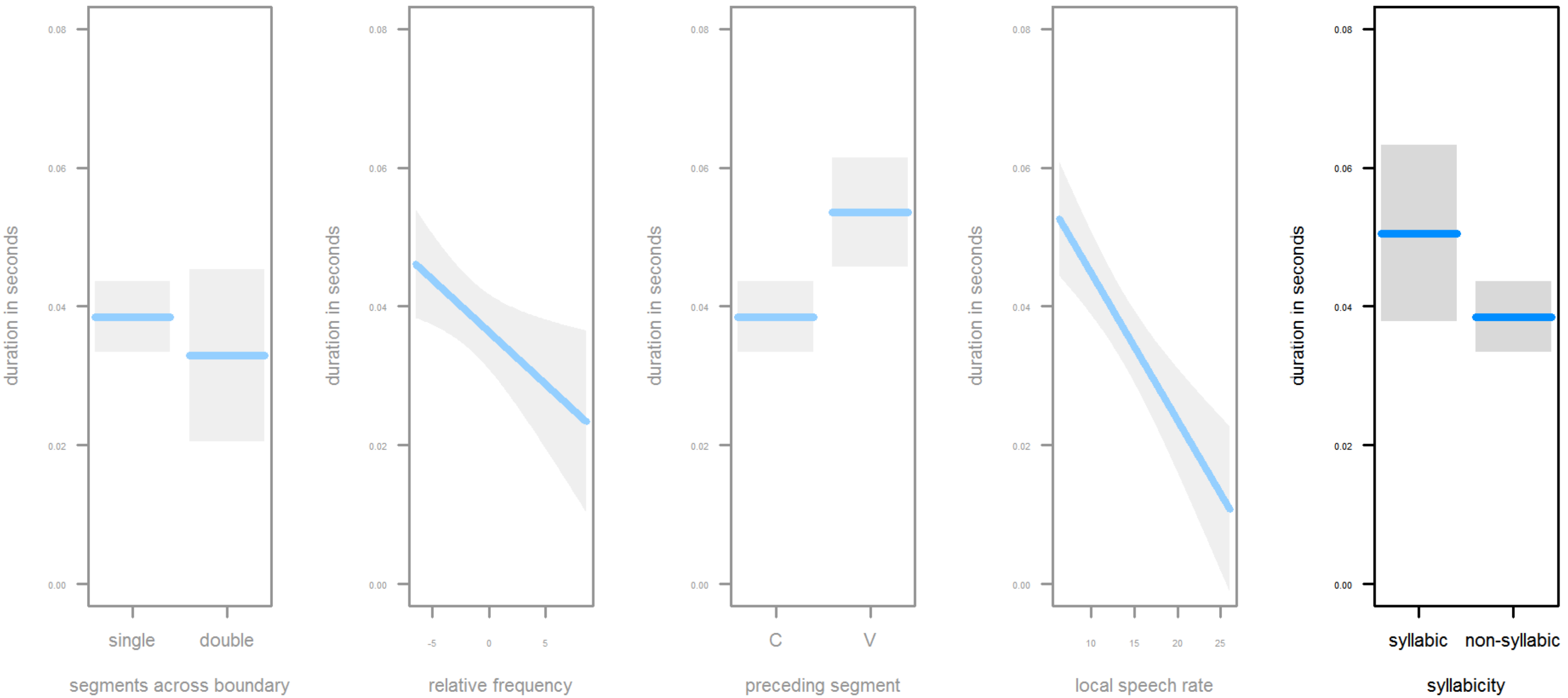
Results 2: *in-* geminates



Results 3: *-ly* does not geminate

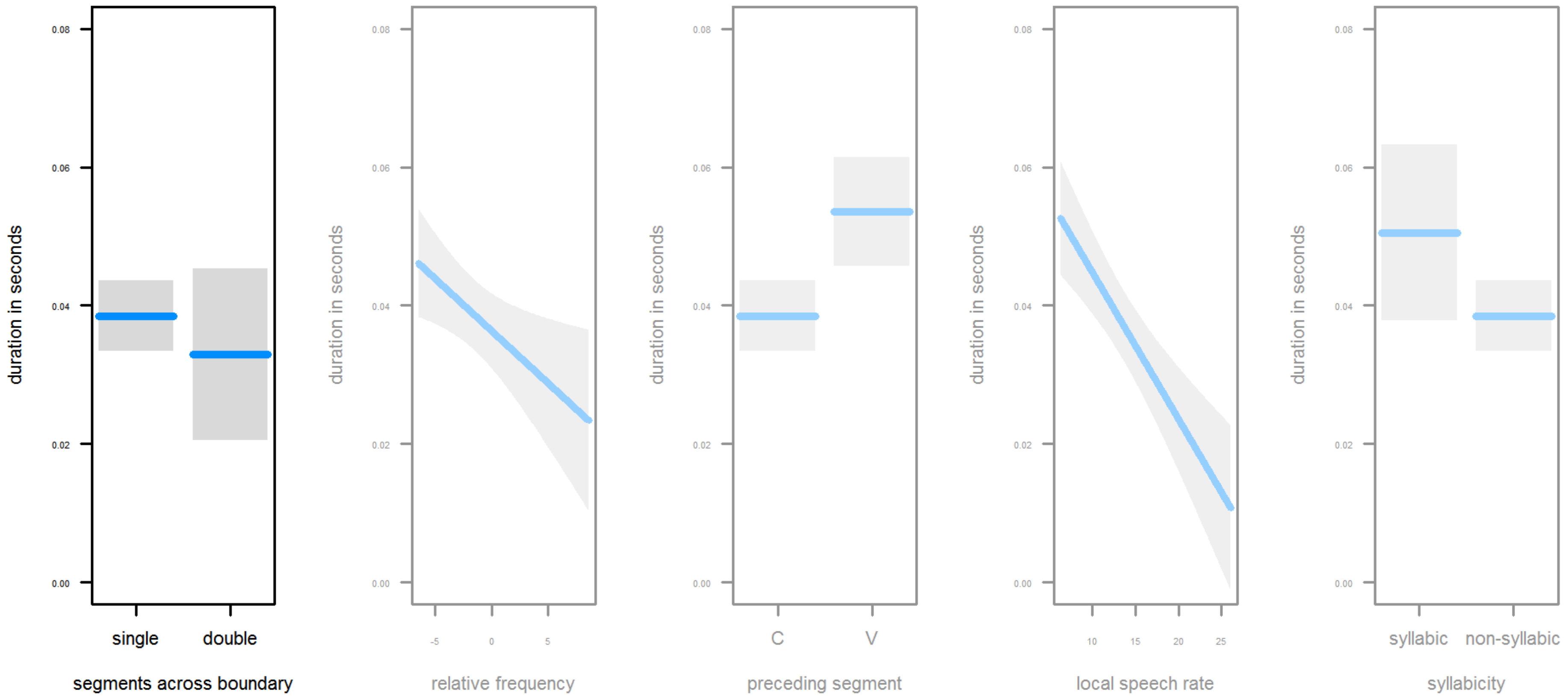


Results 3: *-ly* does not geminate

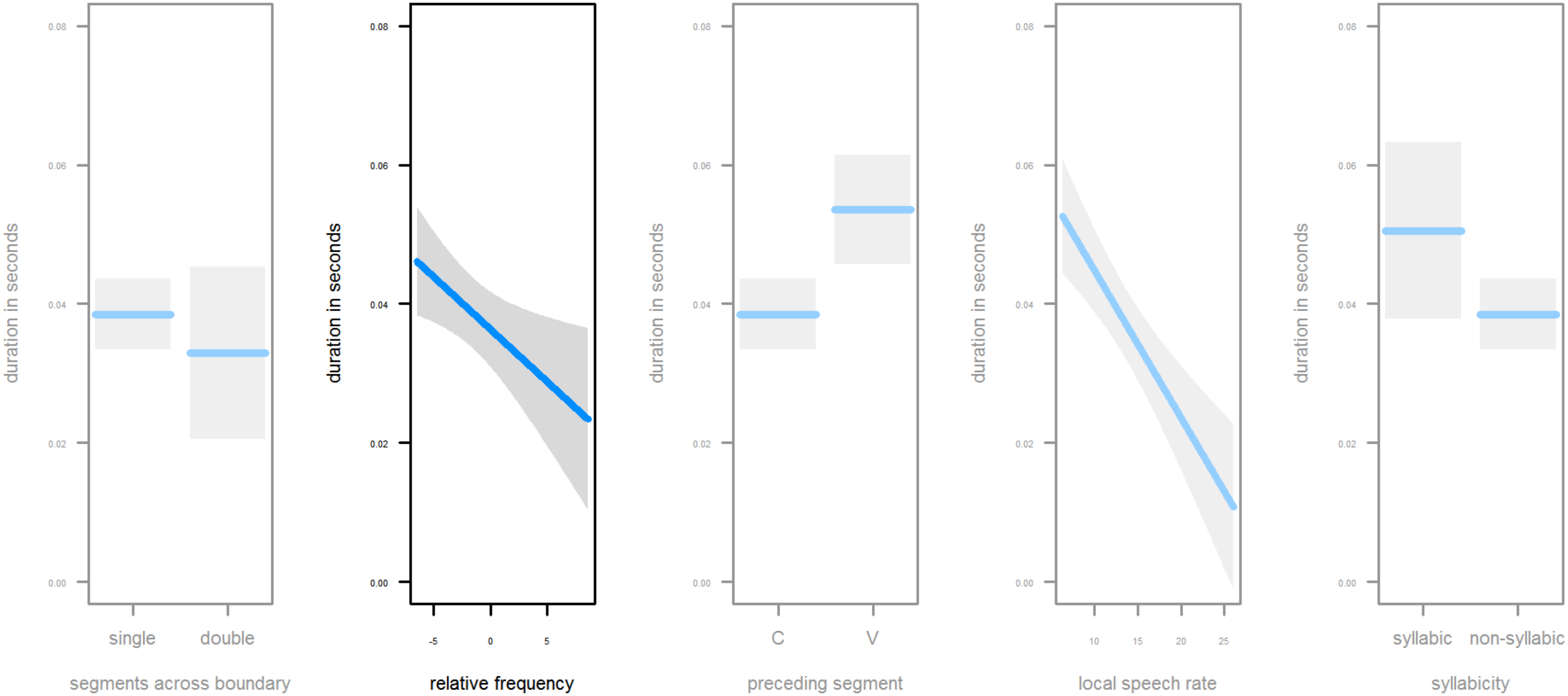


Additional covariate: Syllabicity (ment[l̩]y vs. ment[əl]y, odd[l]y)

Results 3: *-ly* does not geminate



Results 3: *-ly* does not geminate



Summary

- We find morpho-phonological / morpho-phonetic effects, not simple phonetic effects of speech tempo
- *un-* geminates: no surprise
- *in-* geminates: somewhat unexpected result
  - o Effect of AFFIX: homophonous locative and negative *in-* prefixes are acoustically different
- *-ly* degeminates: unexpected result
  - o effect of RELATIVE FREQUENCY: morphological segmentability influences phonetic implementation

Implications

- Empirical facts contradict received wisdom for *in-* and *-ly*
- Lexical Phonology makes wrong empirical predictions
- Morphological information is directly reflected in the speech signal
  - *in-*: Homophonous affixes exhibit different acoustic properties (cf. Plag, Homann & Kunter 2015 on S)
  - *-ly:* Degree of morphological separability correlates with acoustic duration (cf. Hay 2007, Collie 2008)
- Challenges models of lexical phonology and models of speech production that state that post-lexical phonology has no access to morphological information (e.g. Lexical Phonology, Levelt, Roelofs & Meyer 1999)

Thank you very much for your attention!

References

Bauer, L. (2001). *Morphological productivity*. *Cambridge studies in linguistics: Vol. 95*. Cambridge England, New York: Cambridge University Press.

Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxfort reference guide to english morphology*. *Oxford Linguistics*. Oxford: Oxford University Press.

Blazej, L. J., & Cohen-Goldberg, A. M. (2015). Can we hear morphological complexity before words are complex? *Journal of experimental psychology. Human perception and performance, 41*(1), 50–68.

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

Cohen-Goldberg, Ariel M. (2013): Towards a theory of multimorphemic word production: The heterogeneity of processing hypothesis. In: Language and Cognitive Processes 28 (7), S. 1036–1064.

Collie, S. (2008). English stress preservation: the case for ‘fake cyclicity’. *English Language and Linguistics, 12*(03), 505–532.

Cruttenden, Alan; Gimson, Alfred Charles (2014): Gimson's pronunciation of English. 8th ed. London, New York: Routledge.

Giegerich, H. J. (1999). *Lexical Strata in English: Morphological Causes, Phonological Effects*: Cambridge University Press.

Godfrey, John J.; Holliman, Edward (1997): Switchboard-1 Release 2. [Philadelphia, Pa.]: Linguistic Data Consortium.

Hay, J. (2007). The phonetics of 'un'. In J. Munat (Ed.), Studies in functional and structural linguistics: v. 58. Lexical creativity, texts and contexts (pp. 39–57). Amsterdam, Philadelphia: J. Benjamins Pub. Co.

Kaye, A. S. (2005). Gemination in English. *English Today, 21*(2), 43–55.

Kemps, Rachel J J K, Ernestus, M., Schreuder, R., & Baayen, R. H. (2005). Prosodic cues for morphological complexity: the case of Dutch plural nouns. *Memory & cognition, 33*(3), 430–446.

Kiparsky, Paul (1982): Lexical morphology and phonology. In: Linguistics in the morning calm. Selected papers from SICOL-1981. Unter Mitarbeit von The linguistic society of Korea. Seoul, Korea: Hanshin Pub. Co.

Mohanan, K. P. (1986). The theory of lexical phonology. Studies in natural language and linguistic theory: [v. 6]. Dordrecht, Boston, Norwell, MA: D. Reidel Pub. Co.; Sold and distributed in the U.S.A. and Canada by Kluwer Academic.

Oh, Grace E.; Redford, Melissa A. (2012): The production and phonetic representation of fake geminates in English. In: Journal of Phonetics 40 (1), S. 82–91.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Italian Journal of Linguistics / Rivista di Linguistica.*

Plag, I., Homann, J., & Kunter, G. (2015). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*.

R Development Core Team. (2014). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org

un-model

```
# Call:
#   lm(formula = bc ~ TransitionType + LocSpeech, data = unComplex2)
#
# Residuals:
#   Min       1Q    Median       3Q      Max
# -0.081237 -0.027028 -0.000937  0.025328  0.096961
#
# Coefficients:
#   Estimate Std. Error t value Pr(>|t|)
# (Intercept)            0.581989   0.014676  39.655  < 2e-16 ***
#   TransitionTypesingle-C -0.049389   0.009505  -5.196 6.59e-07 ***
#   TransitionTypesingle-V -0.099885   0.009641 -10.360  < 2e-16 ***
#   LocSpeech             -0.007646   0.001063  -7.196 2.83e-11 ***
#   ---
#   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
#
# Residual standard error: 0.03788 on 149 degrees of freedom
# Multiple R-squared:  0.6011,  Adjusted R-squared:  0.5931
# F-statistic: 74.84 on 3 and 149 DF,  p-value: < 2.2e-16
```

im-model

```
#   lm(formula = bc ~ NoCons + LocSpeech + StressPattern + Affix,
#     data = imComplex4)
#
# Residuals:
#   Min       1Q    Median       3Q      Max
# -0.081827 -0.023172 -0.002205  0.023101  0.083318
#
# Coefficients:
#   Estimate Std. Error t value Pr(>|t|)
# (Intercept)            0.2856713  0.0112978  25.286  < 2e-16 ***
#   NoConsdouble          0.0442330  0.0064822   6.824 2.08e-10 ***
#   LocSpeech            -0.0032078  0.0007413  -4.327 2.76e-05 ***
#   StressPatternstr-unstr -0.0344743  0.0071455  -4.825 3.44e-06 ***
#   AffixinNeg            0.0196406  0.0069752   2.816  0.00553 **
#   ---
#   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.0325 on 149 degrees of freedom
# Multiple R-squared:  0.5392,  Adjusted R-squared:  0.5268
# F-statistic: 43.58 on 4 and 149 DF,  p-value: < 2.2e-16
```

-ly-model

```
#   lm(formula = AbsDurCon ~ NoCons + logRelFreq + PrecSegVC + LocSpeech +
#       Syllabic, data = lyComplex2)
#
# Residuals:
#   Min       1Q    Median       3Q      Max
# -0.046194 -0.013208 -0.001831  0.011909  0.045429
#
# Coefficients:
#   Estimate Std. Error t value Pr(>|t|)
# (Intercept)          0.0799558  0.0086899   9.201 3.41e-16 ***
#   NoConsdouble        -0.0074318  0.0056623  -1.313 0.191410
# logRelFreq          -0.0014775  0.0006016  -2.456 0.015219 *
#   PrecSegVCV           0.0168499  0.0047635   3.537 0.000542 ***
#   LocSpeech           -0.0022602  0.0004393  -5.145 8.49e-07 ***
#   Syllabicnon-syllabic -0.0138244  0.0068922  -2.006 0.046726 *
#   ---
#   Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.01876 on 146 degrees of freedom
# Multiple R-squared:  0.2435,  Adjusted R-squared:  0.2176
# F-statistic: 9.398 on 5 and 146 DF,  p-value: 8.768e-08
```

Types

	Doubles	Singles
un-	5	94
in-	17	65
-ly	76	72