

Patterns of /r/ gemination in British and American English: A comparative study

Q. DABOUIS¹, O. GLAIN², S. NAVARRO³

¹UNIVERSITÉ CLERMONT AUVERGNE, LRL (UR 999)

²UNIVERSITÉ JEAN MONNET DE SAINT-ÉTIENNE - UR ECLA

³UNIVERSITÉ DE PARIS - CLILLAC-ARP - UR 3967

Outline

- Gemination in English
- The case of /r/
- Aims
- Methodology
- Results
- Discussion

Gemination in English

Previous studies have shown that English may have “morphological geminates” across morpheme or word boundaries (e.g. *right time, unnecessary*).

Gemination may be seen as a process of phonetic doubling or – more appropriately for English – of phonetic length (Kaye 2005).

These studies have focused on prefixed words with the possible gemination of [l, n, m] (e.g. *illegal, immoral, unnamed*).

Gemination in English

Previous studies have shown that English does have consonantal gemination in semantically compositional prefixed words, although it may vary depending on:

- speech rate
- the productivity of the prefix
- the phonological environment:
 - presence of stress on the following syllable
 - nature of the following segments

(Bauer 2003; Cruttenden 2014: 248; Ben Hedia & Plag 2017; Kaye 2005; Oh 2013; Oh & Redford 2012; Videau 2013).

The case of /r/

None of the previous studies has dealt with the gemination of /r/.

Dabouis (2016) has found that pronunciation dictionaries report a difference between RP and GA regarding the possibility of /r/-gemination:

→ it would be possible to have geminated /r/s in GA, but not in RP in words such as *irrational* or *irremovable*.

The case of /r/

This difference is reported in several dictionaries.

(Kenyon & Knott 1953; Merriam-Webster online; Upton & Kretzschmar 2017; Wells 2008).

Dabouis proposed that it might be attributed to rhoticity:

- A geminated /r/ can be analysed as coda + onset
- Ban on coda /r/ in RP
- No geminated /r/ in RP

There is phonotactic evidence supporting this claim: “degeminated” pronunciations often have light stressed initial pretonic syllables which are otherwise very rare (e.g. *dissatisfy* [ˌdɪsˈsætɪsfai] ~ [(ˌ)dɪ-, -əs-]; *irrational* [ɪˈræʃənəl] ~ [ˌɪ-]).

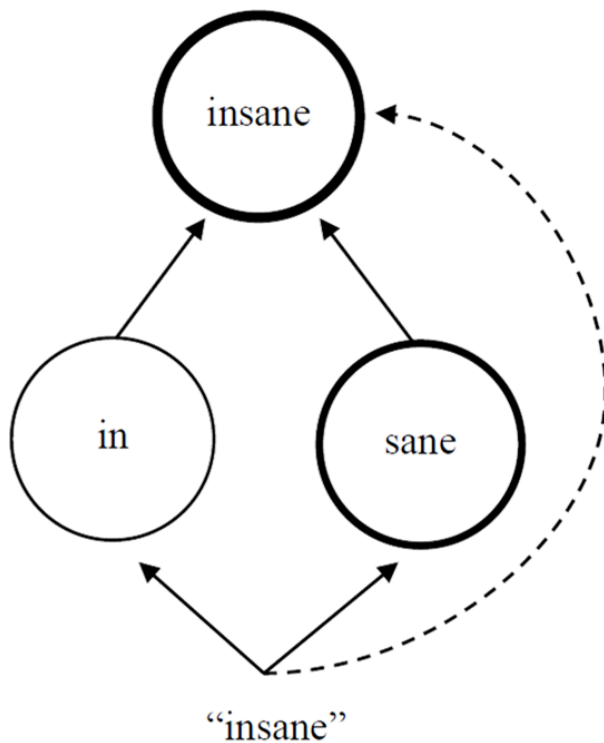
The segmentability hypothesis

Like the previous study by Ben Hedia & Plag (2017), we sought to test out the segmentability hypothesis, which was put forward by Hay (2001, 2003), and states that:

- Complex words which are more segmentable are more likely to preserve the phonological properties of their base
- Complex words which are less segmentable are more likely to drift away from their base phonologically

The segmentability hypothesis

This analysis is based a dual-route race model of lexical access:



-----> Direct route

——> Decomposed route

The route that is adopted depends mainly on the relative frequency of the base and the derivative and on the transparency of their semantic relationship

The segmentability hypothesis

It should be noted that previous studies report mixed results, as certain studies support this hypothesis while others fail to find any segmentability effects (see Plag & Ben Hedia 2018 and Stein & Plag 2021 for overviews)

Aims

Check whether or not the difference between British English and American English can be confirmed using **speech data**.

If attested, evaluate whether the gemination of /r/ depends on the **same variables** as those found for other consonants.

Relevant words

We selected 25 words in <irr-> varying in frequency, semantic transparency, presence or absence of stress on the second syllable and attestedness of the base.

The words are as follows: *irradiate, irradiated, irradiation, irrational, irreconcilable, irredeemable, irrefutable, irregular, irrelevance, irrelevant, irreligious, irreplaceable, irrepressible, irresistible, irrespective, irresponsibility, irresponsible, irretrievable, irreverence, irreverent, irreversible, irrevocably, irrigation, irritate, irritation.*

Youglish-based study

Methodology

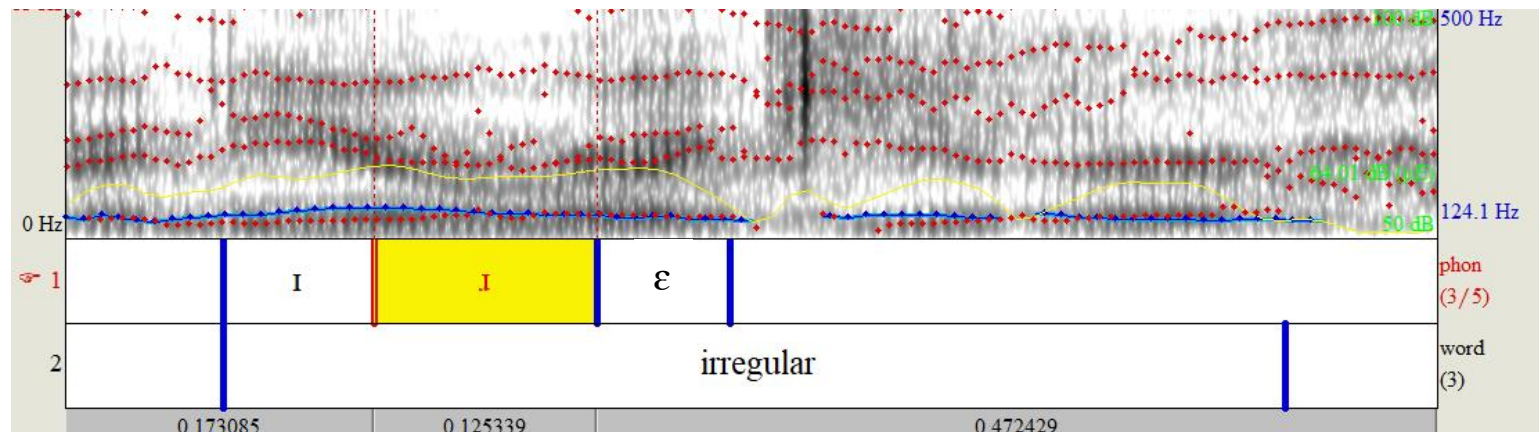
Automatic extraction of the 25 words in Youglish.

A maximum of 25 occurrences per gender per variety were selected.

All words were analysed by two authors

Youghlish-based study

The data was analysed auditorily and spectrographically in Praat (Boersma & Weenik 2014) in order to perform three measurements: the duration of [ɹ], duration of [ɪ] and duration of the whole word.



Youghlish-based study

Dataset

The constitution of the database used in this talk is as follows:

	UK		US		Total
	Males	Females	Males	Females	
Transparent	269	164	454	345	1232
Opaque	51	16	149	98	314
Total	320	180	603	443	1546

Youghlish-based study

Dataset

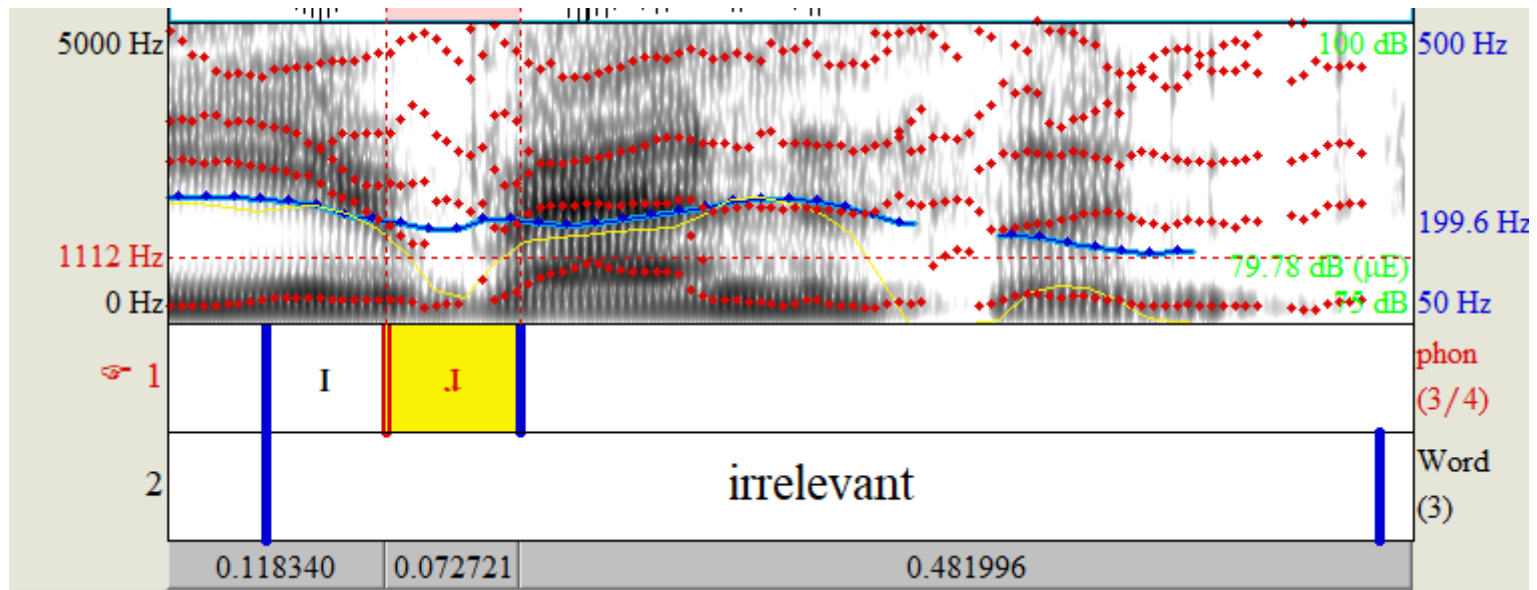
Among the 25 words, 17 are transparent prefixed words (henceforth “derived words”): *irrational*, *irreconcilable*, *irredeemable*, *irrefutable*, *irregular*, *irrelevance*, *irrelevant*, *irreligious*, *irreplaceable*, *irrepressible*, *irresistible*, *irrespective*, *irresponsible*, *irretrievable*, *irreverence*, *irreverent*, *irreversible*.

The remaining eight are:

- 3 with a relatively opaque locative prefix: *irradiate*, *irradiation*, *irradiated*.
- 3 opaque prefixed words: *irrigation*, *irritate*, *irritation*.
- Morphologically ambiguous *irresponsibility*, formed as *irresponsible* + *-ity* or *ir* + *responsibility* and *irrevocably*, formed as *irrevocable* + *ly* or *ir* + *revocably*

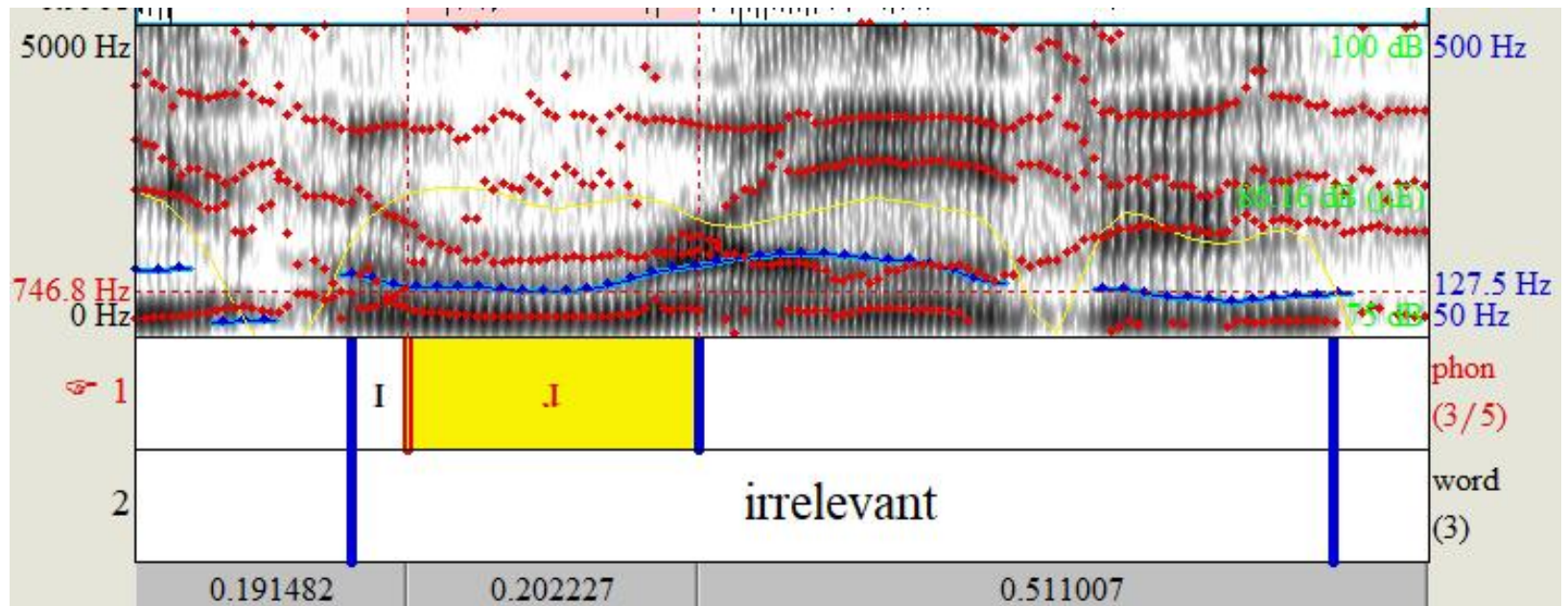
Youghlish-based study

irrelevant, UKF4



Youghlish-based study

irrelevant, USM1



Youghlish-based study

The following variables were coded:

- **R-LENGTH**: Length of [ɹ] in seconds
- **I-LENGTH**: Length of [ɪ] in seconds
- **RATIO OF R LENGTH AND I LENGTH**
- **SPEECHRATE**: Ratio of the number of segments in the word and its length in seconds
- **LOGFREQUENCY**: log-transformed (as $\log_e(x+1)$) frequencies taken from SUBTLEX-UK and US and COCAE
- **LOGFREQUENCYBASE**: log-transformed (as $\log_e(x+1)$) frequencies of bases of derived words taken from SUBTLEX-UK and US and COCAE

Youghlish-based study

- **RELATIVEFREQUENCY:** ratio of the frequency of the derivative and that of base
- **SECONDSYLLABLE:** STRESSED or UNSTRESSED
- **GENDER:** MALE or FEMALE
- **SEMANTICTRANSPARENCY:** the words were coded as TRANSPARENT or OPAQUE
- **ENGVAR:** variety of English, UK or US

Youghish-based study: Results

Statistical analysis: whole dataset

Linear regression was first conducted on the entire corpus, with [ɹ] duration as dependent variable.

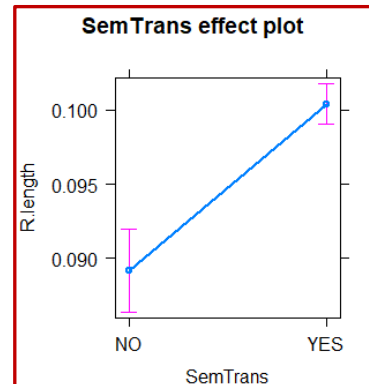
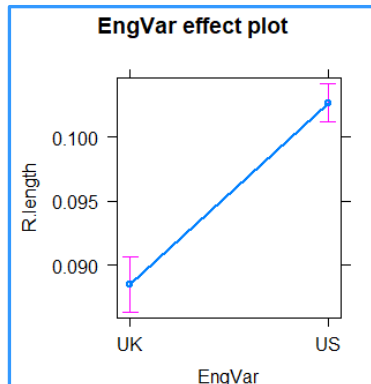
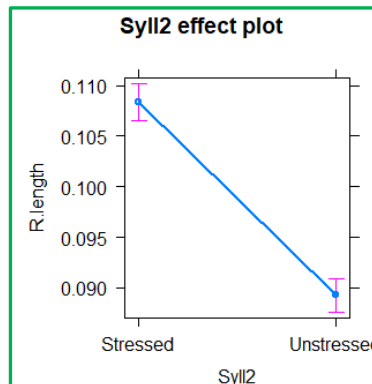
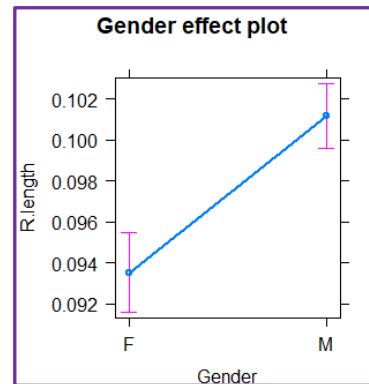
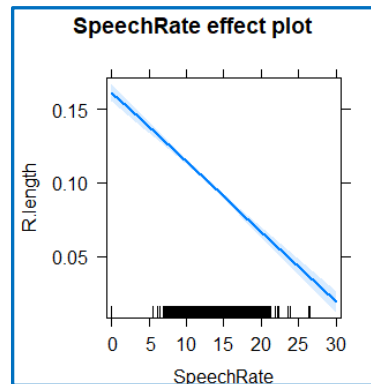
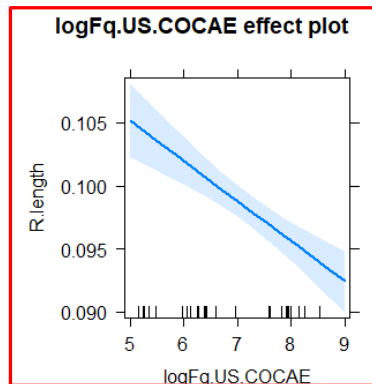
Log-COCAE frequencies were used for both varieties.

All variables were tested and we kept only statistically significant predictors following standard model simplification procedures.

The results are presented with plots of the significant variables.

Youghlish-based study: Results

Both varieties: [ɹ] duration - whole dataset



[ɹ] is longer when
lexical frequency is low

[ɹ] longer if
SpeechRate is low

[ɹ] longer if the
speaker is male

[ɹ] longer if Syll2 is
stressed

[ɹ] longer in US English
than in UK English

[ɹ] longer if the word is
semantically
transparent

Youghlish-based study: Results

Statistical analysis: derived words

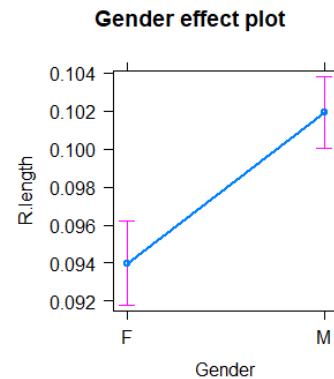
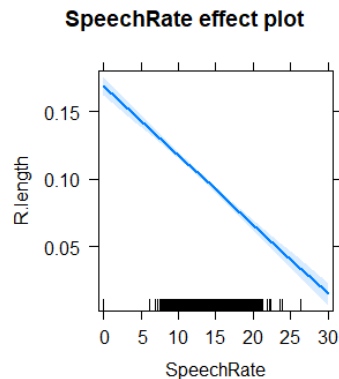
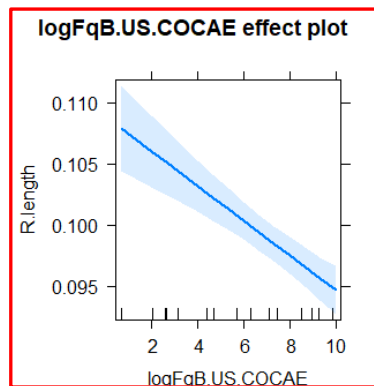
Similar linear regression was then fitted for **derived words** only with variables dealing with the frequency of the base in order to test the segmentability hypothesis:

- **Frequency of the base and frequency of the derivative** as independent variables
- **Relative frequency** (F_{qD}/F_{qB})

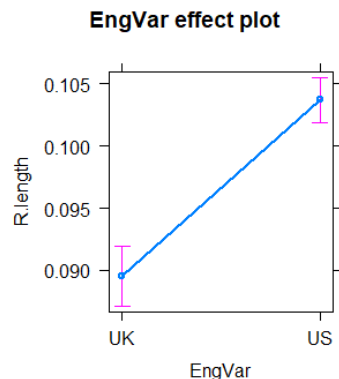
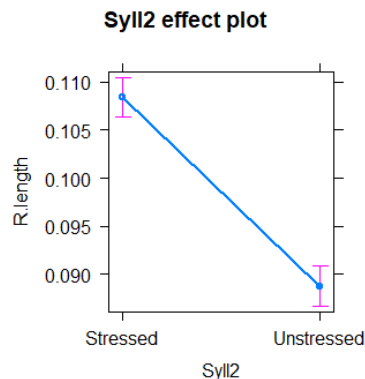
Youghlish-based study: Results

Both varieties: [ɹ] duration – derived words

Base frequency tested through absolute frequency



Frequency of the base neutralizes the effect of frequency of the derivative (probably due to the correlation between the two frequency variables)

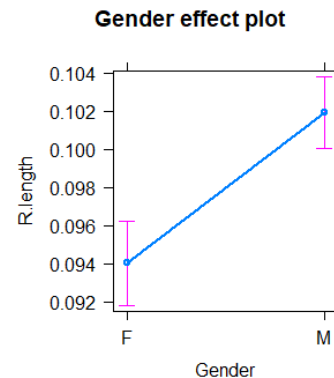
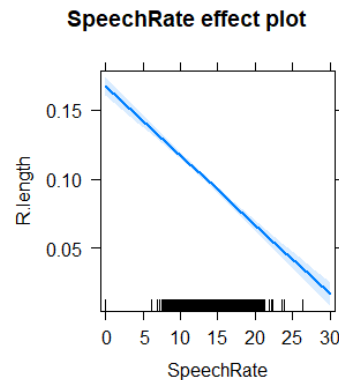
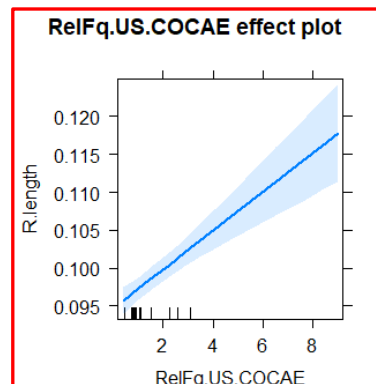


The effect goes in the opposite direction to what is predicted by the segmentability hypothesis!
=> [ɹ] is longer when lexical frequency of the base is low

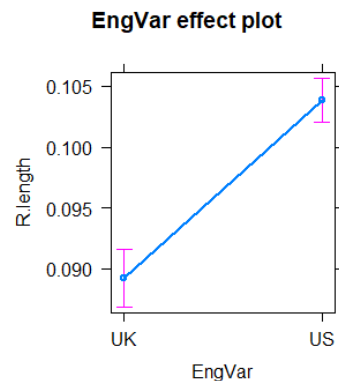
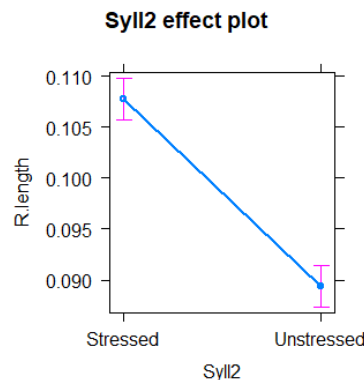
Youghlish-based study: Results

Both varieties: [ɹ] duration – derived words

Base frequency tested through relative frequency



[ɹ] is longer when relative frequency is high (i.e. when derivative is relatively frequent compared with base)



Consistent with the previous model
=> higher frequency of the base tends to shorten [ɹ] duration

Youghlish-based study: Results

UK and US models

Individual models were built for each variety

Overall similar results, except that frequency effects are weaker in UK English

Youghlish-based study: Results

Statistical analysis: [ɪ] duration

Similar tests were run on both varieties with [ɪ] duration as dependent variable. **ENGVAR** was initially included and removed because non-significant.

Whole dataset: Significant effects of SPEECHRATE, SYLL2 and SEMTRANS

- SYLL2: [ɪ] is longer if SYLL2 = UNSTRESSED, => expected as /ɪ/ bears secondary stress in this configuration
- SEMTRANS: [ɪ] is longer in transparent items

Derived words:

- No effect of base frequency whatsoever

Youghlish-based study: Results

Statistical analysis: relative duration of [ʌ] and [ɪ]

Similar tests were run on both varieties with [ʌ]/[ɪ] ratio as dependent variable.

Whole dataset: Similar effects as observed for [ʌ] duration

- Significant effect of word frequency: the higher the frequency, the lower the ratio (i.e. the shorter the [ʌ] compared to the [ɪ])
- Marginal significance of SEMTRANS
- No effect of SPEECHRATE (expected since working with relative duration)

Derived words:

- No effect of base frequency whatsoever

Summary of results

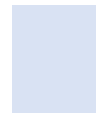
	[ʌ] duration		[ɪ] duration		[ɪ] / [ʌ] ratio	
	Full dataset	Derived words	Full dataset	Derived words	Full dataset	Derived words
LOGFQ	***	***			**	
LOGFQ BASE		***				
RELFQ		***				
SEMTRANS	***		***		*	
SPEECHRATE	***	***	***	***		
SYLL2	***	***	***	***	***	***
GENDER	***	***			***	***
ENGVAR	***	***			***	***



expected
direction



unexpected
direction



no expected
effect



irrelevant

Summary of results

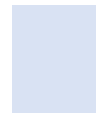
	[ʌ] duration		[ɪ] duration		[ɪ] / [ʌ] ratio	
	Full dataset	Derived words	Full dataset	Derived words	Full dataset	Derived words
LOGFQ	***	***			**	
LOGFQ BASE		***				
RELFQ		***				
SEMTRANS	***		***		*	
SPEECHRATE	***	***	***	***		
SYLL2	***	***	***	***	***	***
GENDER	***	***			***	***
ENGVAR	***	***			***	***



expected
direction



unexpected
direction



no expected
effect



irrelevant

Significant but inconsistent
segmentability effects

Summary of results

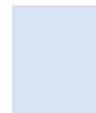
	[ʌ] duration		[ɪ] duration		[ɪ] / [ʌ] ratio	
	Full dataset	Derived words	Full dataset	Derived words	Full dataset	Derived words
LOGFQ	***	***			**	
LOGFQ BASE		***				
RELFQ		***				
SEMTRANS	***		***		*	
SPEECHRATE	***	***	***	***		
SYLL2	***	***	***	***	***	***
GENDER	***	***			***	***
ENGVAR	***	***			***	***



expected
direction



unexpected
direction



no expected
effect



irrelevant

Significant but inconsistent
segmentability effects

No effects of frequency,
gender or variety

Summary of results

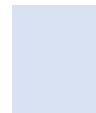
	[ʌ] duration		[ɪ] duration		[ɪ] / [ʌ] ratio	
	Full dataset	Derived words	Full dataset	Derived words	Full dataset	Derived words
LOGFQ	***	***			**	
LOGFQ BASE		***				
RELFQ		***				
SEMTRANS	***		***		*	
SPEECHRATE	***	***	***	***		
SYLL2	***	***	***	***	***	***
GENDER	***	***			***	***
ENGVAR	***	***			***	***



expected
direction



unexpected
direction



no expected
effect



irrelevant

Significant but inconsistent
segmentability effects

No effects of frequency,
gender or variety

No effects of base
frequency

Discussion

Is there a difference between UK and US English with regards to [ɹ] gemination?

Yes, but it is not a categorical difference

- There are perceptively geminated [ɹ]s in UK English

irrelevant, UKM23



The absence of a difference in [ɪ] duration shows that we cannot attribute the higher duration of [ɹ] in US English to an r-coloring effect of the [ɪ] vowel

Discussion

Segmentability

Measured in two ways: SEMTRANS and the base frequency variables (relative & absolute frequency)

- **SEMTRANS:** the effect goes in the expected direction: transparent items have longer [ɹ]s than opaque items
- **Base frequency** (tested among transparent items only): unexpected direction of the effect, as the higher the base frequency, the shorter the [ɹ]

Discussion

Significant base frequency effect for [ɹ] but not [ɪ]

- suggests that segmentability only affects the boundary

In UK English, unclear frequency effects: possibly due to sample size

Overall, only 17 types in derived words → not enough variability in the data to be fully confident in the robustness of the results

Conclusion

There is a difference between UK and US English with regards to [ɹ] duration

This is not a categorical difference, and [ɹ] duration is conditioned by the same variables in both varieties

In both varieties, we find effects of absolute frequency, stress in the second syllable, semantic transparency and gender

We find segmentability effects, but they are contradictory. This is actually consistent with the existing literature

Conclusion

These segmentability effects are circumscribed to the boundary

The limited number of types invites cautiousness regarding frequency effects

- The same variables should be investigated with other consonants

The phenomenon of /r/ gemination and its relationship with segmentability should be investigated further using elicitation tasks with various speech styles, like Oh & Redford (2012) and Oh (2013)

Thank you for your attention!

References

- Bauer, L. (2003) 'The Phonotactics of Some English Morphology', in Jacobsen, H. G. et al. (eds) *Take Danish for Instance. Linguistic studies in honour of Hans Basbøll presented on the occasion of his 60th birthday 12 July 2003*. Odense: University Press of Southern Denmark, pp. 1–8.
- Ben Hedia, S. and Plag, I. (2017) 'Gemination and degemination in English prefixation: Phonetic evidence for morphological organization', *Journal of Phonetics*, 62, pp. 34–49.
- Boersma, P. and Weenink, D. . (2018) 'Praat: doing phonetics by computer [Computer program]. Version 6.1.10'.
- Cruttenden, A. (2014) *Gimson's Pronunciation of English*. 8th editio. Oxon & New York: Routledge.
- Corpus of Contemporary American English* [online]. URL: <http://corpus.byu.edu/coca/>
- Dabouis, Q. (2016) *L'accent secondaire en anglais britannique contemporain*. Ph.D. dissertation. University of Tours.
- Dabouis, Q. (to appear) 'English Phonology and the Literate Speaker: Some Implications for Lexical Stress', in Ballier, N. et al. (eds) *English Word Stress: Theories, Data and Variation*.
- Hay, J. (2001) 'Lexical Frequency in Morphology: Is Everything Relative?', *Linguistics*, 28(6), pp. 1041–70.
- Hay, J. (2003) *Causes and Consequences of Word Structure*. London: Routledge.
- Van Heuven, W. V. J. et al. (2014) 'Subtlex-UK: A new and improved word frequency database for British English', *Quarterly Journal of Experimental Psychology*, (67), pp. 1176–1190.
- Kaye, A. S. (2005) 'Gemination in English', *English Today*, 21(2), pp. 43–55.
- Kenyon, J. S. and Knott, T. A. (1953) *A Pronouncing Dictionary of American English*. Springfield, MA: Merriam.
- Navarro, S. (2016). *Le /r/ en anglais : histoire, phonologie et variation*. Dijon : Editions Universitaires de Dijon.
- Oh, G. E. and Redford, M. A. (2012) 'The production and phonetic representation of fake geminates in English', *Journal of Phonetics*, 40(1), pp. 82–91.
- Merriam-Webster* [online]. URL: <https://www.merriam-webster.com/>
- Oh, G.E. (2013), 'Effects of boundary strength on geminate duration in English', *Studies in Phonetics, Phonology and Morphology* 19.3, pp. 457-478.
- Upton, C. and Kretzschmar, W. A. (2017) *The Routledge Dictionary of Pronunciation for Current English, The Routledge Dictionary of Pronunciation for Current English*. Abingdon & New York: Routledge.
- Stein, S. D. and Plag, I. (2021) 'Segmentability effects on the acoustic duration of affixed words in English', paper presented at the Interface of Phonetics conference in Carl von Ossietzky Universität Oldenburg, Germany.
- Videau, N. (2013) *Préfixation et phonologie de l'anglais : Analyse lexicographique, phonétique et acoustique*. Ph.D. dissertation. Université de Poitiers.
- Wells, J. C. (2008) *Longman Pronunciation Dictionary*. 3rd ed. London: Longman.