Syllable-count judgments: relating gestural composition and syllable weight

Anisia Popescu
Université Paris Diderot - Paris 7

1 Introduction

The concepts of *syllable, consonant* and *vowel* are familiar to everybody, including non-linguists. Speakers can distinguish which segments are consonants or vowels and have fairly robust intuitions about the number of syllables in a word. There is however a specific class of monosyllabic words in American English which has been identified as having variable syllable count judgments (SCJ from now on). Lavoie & Cohn 1999 showed that words involving a tense vowel or a diphthong followed by a liquid (*feel, file, fear, fire*) do not get consistent syllable count judgments from native speakers. Some speakers consider them to be monosyllabic, others disyllabic. When given the choice, some speakers even attribute intermediate ($1 < SCJ < 2$) judgments. The authors term this class *sesquisyllables*. In a followup study, S & Cohn 2016 show that acoustic rime duration is correlated to syllable count judgments. Monosyllabic judgments correspond to words with shorter rimes whereas disyllabic judgments are attributed to words with longer rimes. In terms of phonological representation, Lavoie & Cohn 1999 and Cohn (2003) propose that sesquisyllables are *superheavy* syllables. English, like many languages has a weight constraint for monosyllabic words that requires them to be heavy. This is known as the Minimal Word Condition. In Moraic Theory Hyman (1985), Hayes (1989) this requirement translates into monosyllabic words having to be bimoraic. According to Lavoie & Cohn 1999 and Cohn (2003) sesquisyllables would then be trimoraic, with the tense vowel or diphthong accounting for two moras, and the liquid coda consonant accounting for the third mora.

This paper proposes an articulatory explanation for the variable SCJ observed for sesquisyllables. I argue that the specific gestural composition and the timing of the gestures involved in the production of American English coda liquids are the cause of the observed variable intuitions. American English coda liquids (the dark /l/ and the retroflex or 'bunched' /r/) both have a double gesture: a tongue dorsum (lateral) or tongue body (rhotic) gesture followed by a tongue tip gesture. The tongue dorsum (TD) and the tongue body (TB) gestures are considered to be more vocalic in nature, whereas the tongue tip gesture, in both the lateral and the rhotic, is more consonantal. The vocalic gesture occurs earlier in both liquids. For the lateral there is a retraction of the TD in the uvular region before the TT raising (Sproat & Fujimura 1993) and for the rhotic there is a TB retraction in the pharyngeal region followed by a tongue tip retroflexion (retroflex /r/) or lowering ('bunched /r/'). This paper proposes that the earlier occurrence of the vocalic gesture, adjacent to the vowel gestures of the syllable nucleus is the reason behind the $> 1SCJ$ given to sesquisyllables. In moraic terms this would imply that the vocalic gesture of the

liquid receives an additional mora making sesquisyllables trimoraic. Nasal or stop coda consonants do not have a vocalic gesture in their composition, so no mora is attributed to nasal and coda stops. To investigate this claim, British English and German SCJ were compared in an online task.The comparison between the two languages offers insight into the role of the vocalic gestures on two levels. Firstly, British English, like American English has a coda dark /l/. German has a clear /l/ in coda position. The difference between the two lateral allophones resides in the quality and timing of the vocalic gesture. Dark /l/ is produced with tongue dorsum retraction followed by a tongue tip raising. In clear /l/ the tongue dorsum and tongue tip gesture occur concomitantly. Furthermore the tongue dorsum lowers in the case of clear /l/. One could argue that the TD gesture in clear /l/ is a by-product of the TT raising and has no specific gestural target. In contrast dark /l/ has a specific goal for it's tongue dorsum gesture and this target has to be reached. Comparing British English (or American English) to German would thus yield a single parameter comparison. The choice of British English instead of American English has a double goal. Firstly, the two dialects should not differ if the gestural hypothesis is correct. Secondly, British English is a non rhotic dialect of English. The rhotic has a vocalic gesture, but the consonantal gesture is not produced. If rhotics pattern the same as laterals, this would confirm the role that the vocalic gesture plays in speakers' SCJ.

Table 1 shows the predictions for tense vowel or diphthong nucleus tokens for both British English and German. For British English we expect > 1 SCJ for laterals and rhotics but not for nasals or stops. Sesquisyllables should pattern differently than tokens with non-liquid post-vocalic consonants. Predictions for lax vowel nucleus tokens are stable: no > 1 SCJ are expected for either of the two languages.

	lateral		nasal		open		rhotic		stop	
SCJ	SCJ=1	SCJ>1	SCJ=1	SCJ>1	SCJ=1	SCJ>1	SCJ=1	SCJ>1	SCJ=1	SCJ>1
BE	YES	YES	YES	NO	YES	NO	YES	YES	YES	NO
G1	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO
G2	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Table 1: Prediction for possible SCJ answer option (SCJ=1 or SCJ>1) per language: British English (BE) and German (G)

Open syllables should also get the standard monosyllabic judgment. For German there are two possibilities: either none of the tokens get >1 SCJ (prediction G1), or if they do, we should expect a similar percentages for all coda types (prediction G2). No difference should be observed between *long vowel + liquid* and *long vowel + non-liquid* tokens. Open syllable tokens are not expected to get higher than one SCJ.

2 Methodology

An Ibex Farm Drummond (2010) multiple choice experiment was conducted online. Participants were presented with a word and three answer options: *1 syllable, 1.5 syllables and 2 syllables*. The instructions, inspired by S & Cohn 2016, justified a *1.5 syllables* answer by stating that there is no correct answer and that sometimes participants feel some words have syllable counts in between whole numbers. Fur-

thermore it was specified that *1.5 syllables* did not necessarily mean one syllable and a half, but a value in between a monosyllabic and a disyllabic judgment. Participants were instructed not to rely on spelling, but to subvocalize and base their judgment on their pronunciation.

Stimuli were composed of targets, three types of controls and fillers. Target words consisted of CVC words with a tense vowel or diphthong nucleus and a post-vocalic liquid consonant. Controls consisted of both CV and CVC words. There were two types of CVC controls: either with a tense vowel or diphthong followed by a non-liquid consonant or a lax vowel followed by all types of post-vocalic consonants. Fillers consisted of unambiguous disyllabic words. Table 2 exemplifies the stimuli choice for British English. Reaction times for each answer were recorded.

Stimuli	
Targets	Tense vowel/diphthong + liquid BE: *feel, file, fear, fire* G: *viel* [*fiːl*] *'much', Seil* [*zeil*] *'rope', vier* [*fiːʁ*] *'four'*
Controls	(1) open syllables BE: *fee, tie, pie, too, pay* G: *Fee* [*feː*] *'fairy', sah* [*zaː*] *'saw', sei* [*zei*] *'be'* (2) Tense vowel/diphthong + nasal/stop BE: *zoom, stain, look, pike* G: *Sieg* [*ziːk*] *'victory', Heim* [*heim*] *'home', Teig* [*teik*] *'dough'* (3) Lax vowel + lateral/rhotic/nasal/stop BE: *bill, for, bin, pick* G: *Fell* [*fel*] *'fur', Fett* [*fet*] *'fat', Lamm* [*lam*] *'lamb'*
Fillers	disyllabic words BE: *unite, puppy, party, among* G: *lesen* [*le.zen*[*'to read', Nusskern* [*nuskeʁn*] *'walnut meat'*

Table 2: Stimuli for British English.

For British English there were 20 target lateral and 15 rhotic tokens, 11 nasal, 9 stop, 20 open control tokens, and 30 disyllabic fillers. For German there were 20 target lateral and 10 rhotic tokens, 10 nasal,10 stop and 10 open control tokens as well as 22 disyllabic tokens.

40 British English and 84 German speakers participated in the experiment. Ages ranged between 19 and 85 years old. The geographical area where participants grew up was more varied for the British English than for the German participants. A dialect based analysis was not conducted for this study.

3 Results

Results from the online syllable count judgment task confirm the predictions. The results will be presented in two stages. First results pertaining to the effect of *Vowel type* will be presented, followed by results analyzing the effect of *Coda type*.

3.1 Vowel type

The predictions for the effect of vowel type indicated that lax vowel nuclei tokens should not get > 1 SCJ. This is the case for British English but not German. Figure 1 shows the total count of SCJ per Answer option and Vowel type for both languages. Some German native speakers attribute > 1 SCJ (2.37% of answers) even for short vowel tokens (*Fell* [*fel*] *'fur', Fett* [*fet*] *'fat', Lamm* [*lam*] *'lamb'*. British English native speakers attribute higher than one SCJ exclusively to tense vowel/diphthong nucleus tokens.

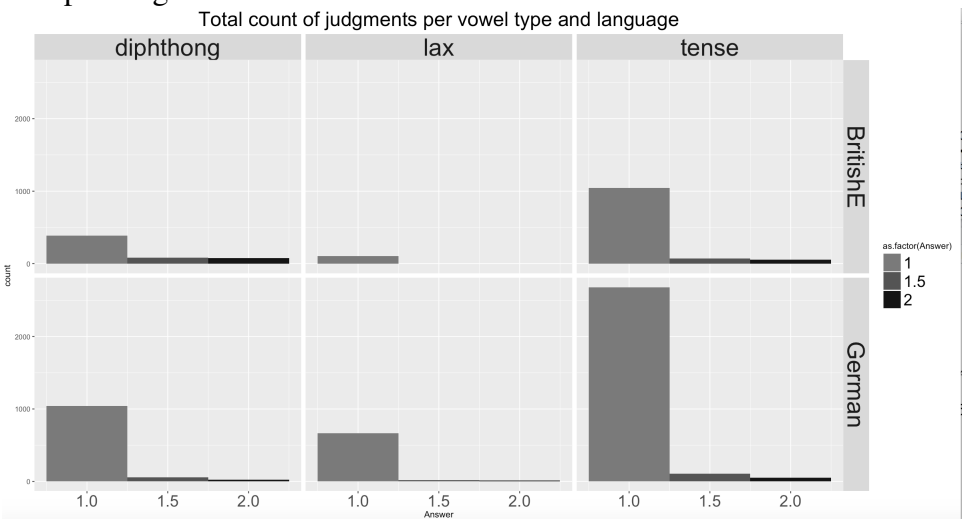

Figure 1: Total count of SCJ per Answer option and Vowel type for British English (above) and German (below)

Figure 2 shows the log values of reaction times per *Language* and *Vowel type*. R (R Core Team, 2018) and *lme4* (Bates *et al.* (2015)) were used to run a linear mixed effects model of the relationship between the interaction of *Vowel type* and *Language* and the logarithmic values of *Reaction times*. The maximal model had *Vowel type* and *Language* as fixed factors and *Speaker* as a random factor. Results show that in the case of British English there is a significant difference between all three vowel types (lax, tense and diphthongs). Diphthong nucleus tokens registered the longest reaction times, with mean values of 3.3 seconds. Reaction times were 0.94 seconds shorter (*t-value = -2,695, p-value=0.007*) for lax vowels and 0.91 seconds shorter for tense vowel tokens (*t-value=-5.266, p-value<0.001*). For German there is a significant difference in reaction time between lax vowels on one hand and diphthong and lax vowels on the other. No difference in reaction times between diphthong and tense vowel tokens was found. SCJ involving lax vowel tokens are attributed ≈ 0.7 faster than SCJ involving tense vowel or diphthong nuclei tokens (*t-value=-2.087, p-vaue=0.03*). Furthermore there is an effect of language German reaction times are overall ≈ 1 second faster than British English response times (*t-value=-4.042, p-value<0.0001*).

3.2 Coda type

The predictions concerning the effect of *Coda type* have also been confirmed. British English participants mark a difference between words involving a tense vowel or diphthong followed by a liquid consonant and words involving a non-liquid consonant. > 1 syllable count judgments are attributed exclusively to sesquisyllables. Figure 3 shows the total count of judgments per answer option for each post-vocalic consonant type.

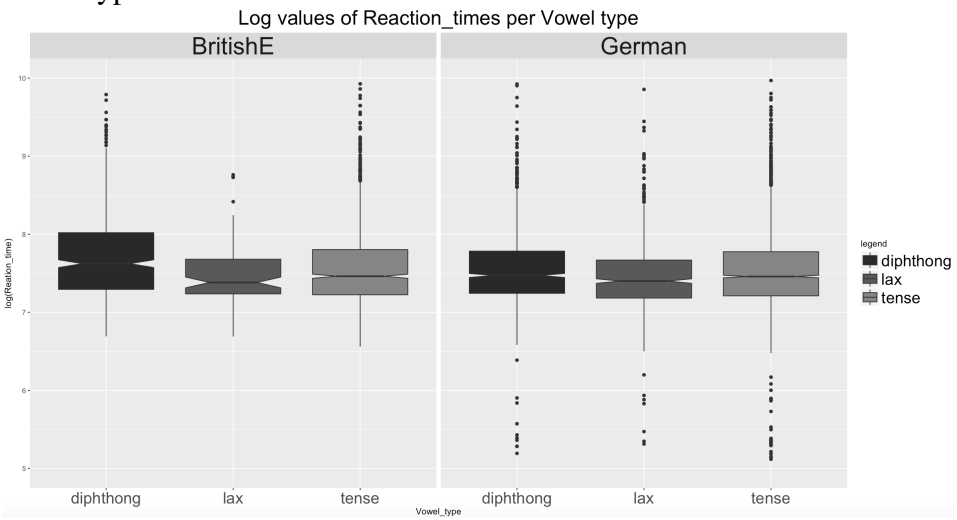

Figure 2: Log values Reaction times per Language and Vowel type

Only answers involving tense vowel or diphthong nucleus tokens are illustrated. Lax vowel tokens were universally considered as monosyllabic. There is however a solitary exception: a *2 syllable* answer is given for the word *meme*. This judgment was awarded by a 85 year old participant who subsequently commented they did not know what the word meant.

In the case of German, > 1 syllable count judgments were attributed to all post-vocalic consonant types. Furthermore open syllables were sometimes considered as disyllabic. Table 3 presents the percentage of attributed judgments for monosyllabic and intermediate (1.5 syllables) or disyllabic options for British English and German for each type of post-vocalic consonant type. Percentages for > 1 SCJ in German for all types of post-vocalic consonants range between 8% and 12%.

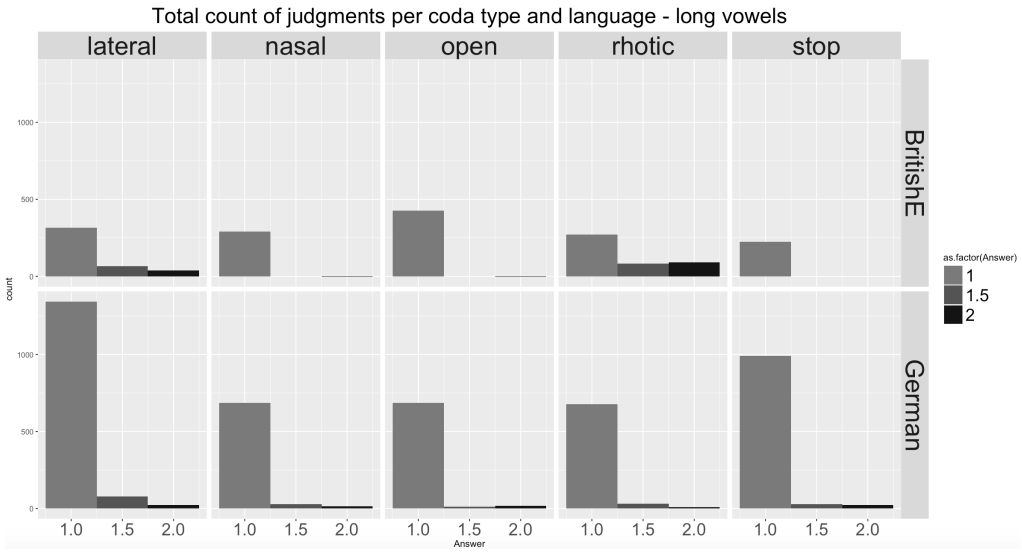

Figure 3: Total count of SCJ per Answer option and Language

This result corresponds to the *G2* prediction in Table 1 which predicts and similar > 1 SCJ percentages will be awarded to all tokens, independent of coda consonant type. The instructions and the three answer choice option including the *1.5 syllables* option had a priming effect. When given the choice participants opted for the *1.5 syllables* answer. The lack of difference between the percentages of *1.5 answers* however indicated there is no discernible pattern, contrary to the British English data, for which there is a clear pattern: *1.5 syllables* SCJ were attributed exclusively to tense vowel/diphthong + liquid consonant tokens. These results indicate that sesquisyllables are a special class of monosyllabic words *only* in British English and not German.

SCJ	lateral		nasal		open		rhotic		stop	
Language	BE	G	BE	G	BE	G	BE	G	BE	G
1	85%	88%	100%	90%	100%	98%	70%	92%	100%	90%
> 1	15%	12%	0%	10%	0%	2%	30%	8%	0%	10%

Table 3: Percentages of attributed judgments for *1 syllable* and *1.5 or 2 syllables* for British English (BE) and German (G) for each type of post-vocalic consonant

Results show an unexpected difference in SCJ attribution between lateral and rhotic coda tokens in British English: there are twice as many > 1 SCJ attributed to rhotic coda token than lateral coda tokens.No specific predictions were made about the ratio of > 1 SCJ but a comparable ratio was expected for both types of liquid coda tokens. A by-token analysis offers a possible explanation. For the lateral coda targets there are two categories of tense vowel tokens: back vowel + lateral (*fool, pool, tool*) and front vowel + lateral (*feel, teal, meal*). SCJ for the two categories differ: no > 1 SCJ are given to *back vowel+lateral* tokens whereas *front vowel + lateral* tokens get attributed 15% of > 1 SCJ. To a smaller degree the opposite is found for German: German participants tend to attribute more > 1

SCJ to tokens involving *back vowel + lateral*. If we take into account that clear /l/ resembles the articulation of /i/ while dark /l/ resembles that of /u/ (Sproat & Fujumura 1993) this result is not surprising. There are two points to be made. First, we expect an increased gestural overlap (blending) between similar consecutive articulatory gestures (German *Stiel*, English *stool*). If there is gestural overlap, one might argue that only one weight unit is associated with both gestures. Second, different consecutive gestures (German *Stuhl*, English *steel*) trigger coarticulation, rendering the English dark /l/ clearer, in the sense that the tongue dorsum gesture is reduced) and the German clear /l/ darker, thus explaining lack of > 1 SCJ for English *front vowel + lateral* and the higher amount of > 1 SCJ for German *back vowel + lateral*. Both explanations point to a possible association between weight units and vocalic gestures.

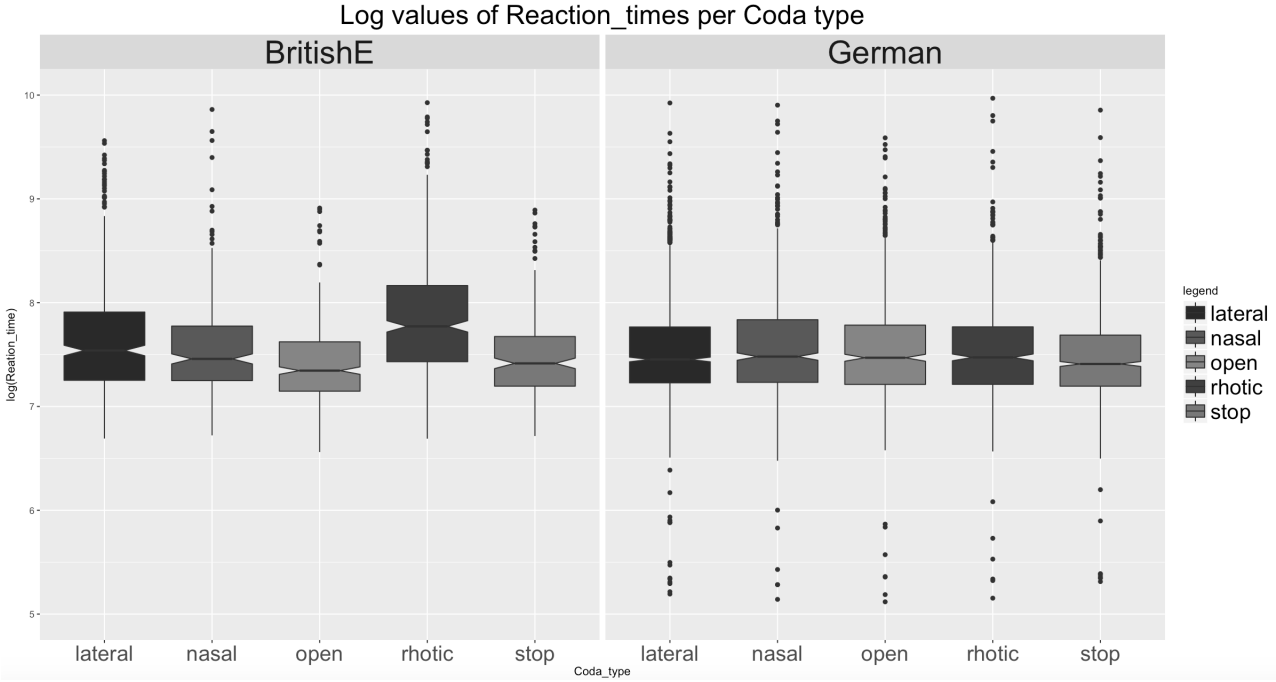

Figure 4: Log values Reaction times per Language and Coda type

For each answer, reaction times were recorded. Reaction times varied from 0.7 to 15 seconds. Figure 2 shows the logarithmic values of reaction times for each language per coda type. A maximal linear mixed effects model with *Coda type* and *Language* as fixed factors and *Speaker* as a random factor was carried out. Results show that for British English there is a coda type effect. Participants took a shorter amount of time to attribute syllable count judgments for tokens with nasal and stop post-vocalic consonants. Mean response times for lateral tokens were of ≈ 7 seconds. Response times were 0.15 seconds shorter (*t-value = -2.535, p-value= 0.014* for nasal tokens and 0.23 seconds shorter (*t-value = -3.911, p-value= 0.0002*) for stops. In the case of rhotics, participants took an overall longer time to attribute a SCJ, ≈ 0.15 seconds longer (*t-value=3.026, p-value = 0.004*). German participants took the longest time to attribute SCJ to nasals (≈ 7.6 seconds). Compared

to nasals, reaction times for liquids are in between 0.06 (laterals:*t-value = -2.417, p-value = 0.01*) and 0.1 seconds shorter (rhotics: *t-value = -3.47, p-value = 0.009*). Tokens with post-vocalic stops take the shortest amount of time to analyze: mean values are of 7.5 seconds.

Reactions times offer an insight into participants' processing of syllable count judgments. Even if the participants do not choose > 1 SCJ for theoretically monosyllabic words, the higher reaction times for target tokens involving coda liquids show that sesquisyllables form a special class of monosyllabic words. For German, on the other hand, tokens with long vowels followed by a nasal seem to be the hardest words to process. Liquids pattern with open syllables showing that sesquisyllables do not exist in German.

4 Discussion

This study examined the relationship between gestural composition of liquid coda consonants and speakers' intuitions about syllable count judgments. The aim was to propose a gestural account for the existence of a special class of American English monosyllabic words identified as *sesquisyllables* byLavoie & Cohn 1999. The hypothesis was that the presence and the timing of the vocalic gesture involved in the production of American English coda liquid consonants is the cause of speakers' variable intuitions. To test this hypothesis an online task was carried out comparing British English and German participants' syllable count judgments. The choice of languages was motivated by the distribution of coda liquid allophones.British English allowed an investigation of the role of the vocalic gesture involved in the rhotic coda consonant by comparing the present results with a previous study done on American English. The comparison between British English and German allowed for a direct comparison between coda lateral consonants. The results confirmed the predictions made based only on the gestural composition of the coda liquids were confirmed. The study confirmed the presence of sesquisyllables in a different variety of English (British English), and showed that German, a language that does not have a separately timed vocalic gesture in the composition of its coda liquids, does not have sesquisyllables.

These results suggest that syllable weight is linked to gestural composition. In order to further prove this claim a paired production and syllable count judgment should be carried out, looking for correlations between articulatory production and intuitions about syllable count.

References

Bates, B., M. Maechler, B. Bolker, & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48.

Cohn, A. 2003. Phonological structure and phonetic duration: The role of the mora. *Working Papers of the Cornell Phonetics Laboratory* 15.69–100.

Drummond, A. 2010. Internet based experiments (ibex)(version 0.3)). `http://spellout.net/ibexfarm`. Retrieved 12 February 2015.

Hayes, B. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry 20* 20.253–306.
Hyman, L. 1985. *A Theory of Phonological Weight*. Dordrecht: Foris.
Lavoie, L., & A. Cohn. 1999. Sesquisyllables of enlgish: The structure of vowel-liquid syllables. In *Proc. 14th ICPhS Berkeley*, 109–112.
S, T., & A. Cohn. 2016. Shared representations underlie metaphonological judgments and speech motor control. *Laboratory Phonology* 7.1–33.