

How do word frequency and syllable surprisal affect response time and acoustic duration in sentence formulation?

Ivan Yuen, Bernd Möbius, Bistra Andreeva, Mitko Sabev

Language Science and Technology Saarland University, Germany

{ivyuen, moebius, andreeva, msabev}@lst.uni-saarland.de

Abstract

Although predictability effects have been shown to affect acoustic realization (e.g., [1 - 8]) at different linguistic levels, they are mostly based on variation of predictability at a specific linguistic level. Yet linguistic levels are embedded within one another. ‘Frequency of occurrence’ is one type of predictability, which could occur at the level of word (word frequency) and syllable (syllable frequency). The current study explored such cross-level interaction on response time and acoustic vowel duration in German, in the context of the psycholinguistic model of speech production in [14]. We hypothesized an additive effect; however, our findings revealed selective interaction, depending on whether the word is mono- or disyllabic. These results challenge the assumption of ‘discrete, staged and serial’ processing across levels in [14].

Index Terms: speech production, word frequency, syllable surprisal, response time, acoustic vowel duration

1. Introduction

Linguistic predictability is pervasive and one type of predictability is related to exposure and usage: occurrence frequency. The ‘frequency’ effect has been reported to occur at multiple linguistic levels, such as word and syllable. It has been shown to affect speech initiation response time (RT) and various acoustic measures with fast RT, short duration and reduced spectral characteristics for frequent units [1 - 13]. However, those studies tend to focus on either RT or acoustic measures at one specific linguistic level. It is not clear how ‘frequency’ might play out across multiple linguistic levels, for example, word and syllable, and whether RT and acoustic measures characterize the same kind of process in speech production planning.

At the lexical level, high frequency (HF) words are easy to access, and therefore fast retrieval leads to short RT [11]. Similarly, HF words are reported to exhibit short acoustic duration [3 - 5, 7, 13]. From an information-theoretic perspective, speakers economize their effort in producing HF words which are predictable from usage, without compromising lexical access.

Likewise, frequency at the syllable level is found to have fast access leading to short RT [9 - 10, 12] and exhibit short acoustic duration [8]. According to the psycholinguistic model of speech production [14], the syllable frequency effect on RT has been argued to be a post-lexical process, arising from the retrieval in the mental syllabary [9, 12, 14]. Short RT arises, because the high frequency syllable is stored and precompiled with articulatory specification from the phonological code (i.e. syllabification). Precompilation can save the time needed to compute online the articulatory specification, implying a tight relationship between RT and phonetic encoding. Consequently, the effect of syllable frequency on acoustic duration emerges from the retrieval of a stored and precompiled high frequency syllable. However, the previous studies mostly manipulated syllable frequency in pseudo-words with some exceptions such as [12] in which word frequency covaries with syllable frequency.

But real words can have high or low frequency, each of which can vary in syllable frequency as well. This then raises

the question of how word frequency and syllable frequency could affect RT and acoustic duration, which the current study intends to explore. The current study indexed syllable frequency in terms of ‘surprisal’. A high syllable surprisal suggests low syllable frequency and low predictability, whereas a low syllable surprisal suggests high frequency and high predictability; see section 2.1 for a formal definition of surprisal.

According to the model in [14], different processing stages are assumed to be discrete, staged and serial. That is, the word frequency effect (i.e., any advantage in RT from the high frequency word) at the lexical level will be percolated downstream to lower levels for phonological and phonetic encoding. In light of that, we hypothesize short RT for high frequency words and long RT for low frequency words, with an additive effect of syllable surprisal (short RT for low surprisal syllables and long RT for high surprisal syllables). We also hypothesize short acoustic duration for low surprisal syllables because of the precompiled phonetic form. Given the ‘discrete and serial’ processing assumption of the model in [14], we expect the two hypotheses to hold for both monosyllabic and disyllabic words. The predictions are schematized in Figure 1.

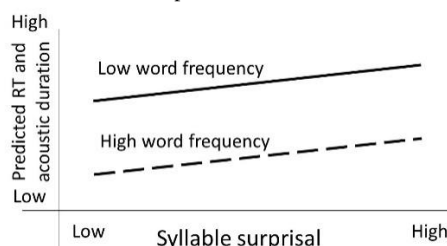


Figure 1: Predicted effects of word frequency and syllable surprisal on predicted response time (RT) and acoustic duration.

Table 1: High frequency and low frequency monosyllabic and disyllabic word stimuli, with log syllable surprisal values on the stressed syllable.

Syllable type	High freq. word/Syl. surprisal	Low freq. word/Syl. surprisal	Stressed Vowel
Monosyl.	Kind (child)/17	Gift (poison)/17	/ɪ/
	Mensch (human)/12	Mett (mince)/17	/ɛ/
	Zahl (number)/1	Tal (valley)/13	/a:/
	3		
	Tag (day)/25	Zahn (tooth)/15	/a:/

Disyl.	Katze (cat)/13	Tatze (paw)/13	/a/
	Zimmer (room)/14	Gitter (grid)/13	/ɪ/
	Fehler (error)/13	Feder (feather)/13	/e:/
	Meter (meter)/10	Feta (feta)/13	/e:/
	Vater (father)/10	Faser (fibre)/10	/a:/

2. Method

2.1. Stimuli and participants

The stimuli consisted of monosyllabic and disyllabic German words, in which occurrence frequency of words and syllables were varied. While a syllable is the same as a word in the former, a syllable forms part of a word in the latter. By including both monosyllabic and disyllabic words, we aimed to separate any effect of word frequency from that of syllable surprisal. The choice of stimuli was subject to multiple constraints. The final stimuli were chosen to be as different as possible in word frequency with a range of syllable surprisal, while keeping the syllable structure, consonantal voicing and vowel identity as similar as possible. An additional constraint was that these words can be depicted in a picture. The selected stimuli (Table 1) were also varied to occur in two utterance positions: medial, e.g., *Das rechte Gift ist grün* (gloss: The right-side poison is green) vs. final, e.g., *Der Mann trinkt das Gift* (gloss: The man drinks the poison), to examine further whether the frequency effects at both word and syllable levels can be generalized across utterance positions. The two utterance positions were elicited by two auditory prompt questions respectively: *Welches Gift ist grün?* (Which poison is green) for the former; *Was trinkt der Mann?* (What does the man drink) for the latter. The stressed vowel of the stimuli was matched as closely as possible in different word frequency/syllable surprisal manipulations, restricted broadly to a small set of ‘tense/long’ and ‘lax/short’ vowels.

Word frequency was estimated using SUBTLEX-DE [15] and CELEX [16], whichever has a higher count. Binning of high (HF) vs. low (LF) word frequency was based on a threshold of 1000/million. High frequency words ranged from 1100 to 16542/million, and low frequency words from 0 to 630/million. The word frequency measure was consistent with word-based unigram surprisal.

Syllable-based unigram surprisal, defined as $-\log_2 P_{\text{Probability}}(X)$, was estimated from the best-performing language model trained on the deWaC (.de domain Web as Corpus) [17]. The corpus is a collection of web-crawled data containing about 1.7 billion word tokens and 8 million word types from diverse genres. The corpus was first pre-processed and normalized by removing unnecessary/irrelevant/duplicate document information, using German Festival TTS [18]. The normalized corpus was then split into a training set (80%) and a test set (20%). Language models were trained using the SRILM toolkit [19] and underwent Witten-Bell smoothing [20]. The test set was used to identify the best-performing language model.

Syllable surprisal ranged from 10 to 25 for the high frequency words and from 10 to 17 for the low frequency words. The average syllable surprisal was higher in the high frequency (17) than the low frequency (16) monosyllabic words. The direction was reversed in the disyllabic words: the average syllable surprisal value was lower in the high frequency (12) than low frequency (16) words. Note that the syllable surprisal estimates were applied to the first stressed syllable in

the disyllabic words after the application of the onset maximization principle. Figure 2 shows the relationship between word frequency and syllable surprisal of the stimuli.

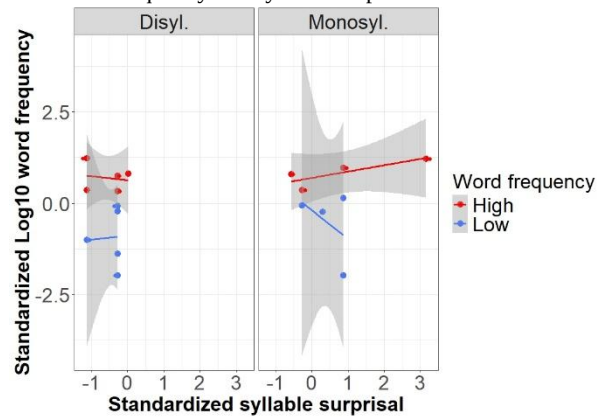


Figure 2: Relationship of standardized word frequency and syllable surprisal values of the monosyllabic and disyllabic stimuli.

Twenty participants took part (9M, 11F, mean age = 23.4 years, range = 19-42 years) and were remunerated for their participation, with informed consent in accordance with the approved ethics guideline at Saarland University.

2.2. Procedures

The production experiment consisted of two parts: a picture-naming task followed by a verbal sentence response generation task (i.e. test phase). Prior to testing, two practice items were used to familiarize participants with the task. In the picture-naming task, participants were instructed to name a pictured object, with the goal of ensuring that the intended target word associated with the corresponding object would be used in the verbal sentence response generation task in the test phase. Testing only began when they succeeded in associating the intended target word with the corresponding picture in the naming task. In the test phase, they were presented with the picture of a target word to generate a verbal sentence response to an auditory question as soon as possible. The test phase contained two blocks, one for the utterance-medial stimuli, and another for the utterance-final stimuli. Each trial began with a fixation cross (1 second), followed by a visual display (3 seconds) for previewing. After previewing, participants clicked a button to hear one of the auditory prompt questions and were instructed to respond as fast as possible. The two blocks were repeated. The experiment was created on Labvanced¹ for the presentation of visual stimuli, auditory prompt questions, as well as randomization. The prompt questions and their sentence responses were recorded to a desktop computer using Audacity² at a sampling rate of 44.1 kHz.

2.3. Annotation

The target word and stressed vowel of the word were annotated in Praat [21], with reference to waveform and spectrographic information. Word level annotation was guided by specific acoustic landmarks associated with different consonant types in syllable onset or coda position. The onset of burst release was used to identify the beginning of words with an onset plosive consonant. The onset of frication was used to identify those with an onset fricative consonant. The onset of voicing and nasal formants were used to identify those with an onset nasal consonant. The offset of a disyllabic word was identified as the

¹ <https://www.labvanced.com/>

² <https://www.audacityteam.org/>

end of F2, because it always ends with an unstressed vowel. Since a monosyllabic word ends with a coda consonant, its word boundary was identified according to consonant types. For coda plosive consonants, the end of release/aspiration was used. For coda fricatives, the end of frication was used. For coda nasals, the end of nasal formants was used. For laterals, the end of F2 was used.

At the vowel level, the beginning of target vowels was identified in terms of clear sharp F1 and F2 onset, and the end of target vowels in terms of clear F2 offset. The annotated vowel boundaries took into consideration additional acoustic landmarks from the immediately adjacent consonant in the onset and coda positions, for example, frication noise for onset fricatives, the closure, release/aspiration for plosives, nasal formants for nasals, the closure/release phases for affricates, or the sharp turning point in F2 transition/lateral release for laterals. Additional acoustic cues were used to identify laterals, such as a change in the amplitude of a waveform arising from the secondary resonance chamber.

Speech initiation response time (RT) was measured from the onset of the prompt questions to the onset of the sentence response. Vowel duration was automatically extracted from the annotation in Praat.

2.4. Analysis

Prior to statistical analysis of 1335 items using a linear mixed effects model and lmerTest in R [22 - 24], speech initiation response time (RT) and vowel duration of the stressed syllable were log10-transformed to adjust for skewness. Syllable surprisal was standardized and centered. Linear mixed effects models were constructed for RT and vowel duration, testing for the effects of WORD FREQUENCY, SYLLABLE SURPRISAL, SYLLABLE, and POSITION as fixed predictors, specifically the interaction between WORD FREQUENCY and SYLLABLE SURPRISAL. SPEAKER and WORD were included in the random structure with intercept and by-factor slope, unless their inclusion resulted in overfitting or failure in model convergence. Optimal model fit was then identified using AIC and likelihood ratio tests. Significance testing was carried out using the anova() function, with Satterthwaite approximation to estimate the degree of freedom.

3. Results

3.1. Speech initiation response time (RT)

The structure of the final model using log10RT as a dependent variable was: $\sim$ SYLLABLE SURPRISAL * WORD FREQUENCY * SYLLABLE * POSITION + (SYLLABLE + POSITION | SPEAKER) + (1 | WORD). The results yielded a significant main effect of syllable surprisal ($F(1, 16.89) = 9, p = .006$), a significant 2-way interaction of syllable surprisal-by-position ($F(1, 1265.64) = 4.9, p < .001$), and a 4-way SYLLABLE SURPRISAL-BY WORD FREQUENCY-BY-SYLLABLE-BY-POSITION interaction ($F(1, 1266.1) = 5.1, p = .02$). To better understand the 4-way interaction, we carried out further sub-analyses of monosyllabic and disyllabic words separately.

3.1.1. Monosyllabic words

The optimal model for the monosyllabic words had the structure of $\sim$ SYLLABLE SURPRISAL + WORD FREQUENCY + POSITION + (POSITION + SYLLABLE SURPRISAL | SPEAKER) + (1 | WORD). As the full model with interactions reached singularity, a reduced model with syllable surprisal/word frequency-by-position interactions was used to compare to the optimal model. Based on the likelihood ratio test, the additional interactions did not result in any improvement of the model fit: $\chi^2(2) = 1.68, p = .43$. The final model yielded significant main effects of WORD FREQUENCY ($F(1, 7.55) = 7.55, p = .026$), and POSITION ($F(1, 20.43) = 4.82, p = .04$) (Figure 3). Overall, low frequency (LF) words are responded to faster than high frequency (HF) words,

while it takes longer to respond in medial than final utterance position.

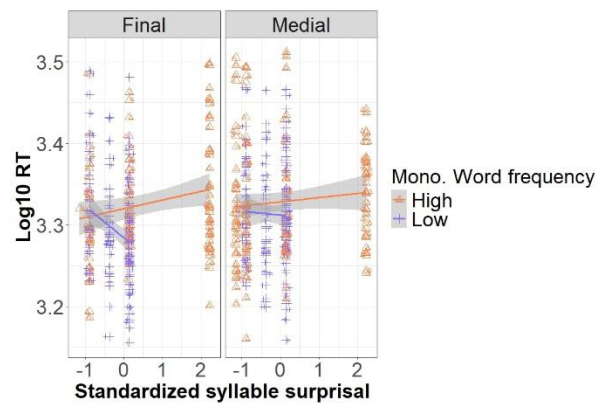


Figure 3: RT (log) as a function of standardized syllable surprisal in high- and low-frequency monosyllabic words in medial and final utterance positions.

3.1.2. Disyllabic words

The optimal model for the disyllabic words had the structure of $\sim$ SYLLABLE SURPRISAL + WORD FREQUENCY + POSITION + (POSITION | SPEAKER) + (POSITION | WORD). Based on the likelihood ratio test, the full model with interactions did not improve the model fit: ($\chi^2(4) = 2.2, p = .7$). The optimal model yielded a marginally significant main effect of SYLLABLE SURPRISAL ($F(1, 9.3) = 4.3, p = .06$) (Figure 4), suggesting that speech may be initiated earlier for a less frequently occurring syllable than a more frequently-occurring one.

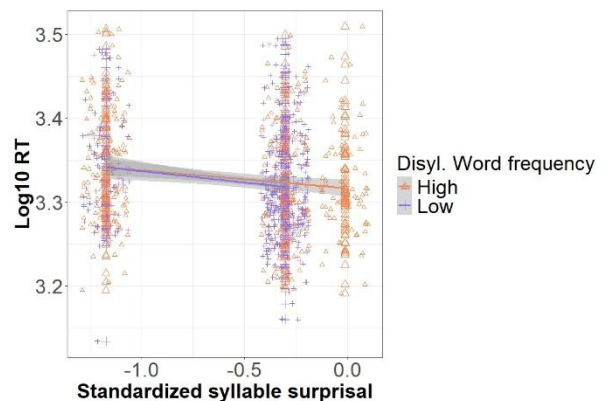


Figure 4: RT (log) as a function of standardized syllable surprisal in high- and low-frequency disyllabic words.

In other words, the effects of WORD FREQUENCY and POSITION affect RT to the monosyllabic words; whereas SYLLABLE SURPRISAL influences that to the disyllabic words. We expected high frequency words to have short rather than long RT, but this was not the case. We also expected RT to increase as syllable surprisal increases. While this was the case for the high frequency monosyllabic words, this was not for the low frequency monosyllabic words (Figure 3). As regards the disyllabic items, syllable surprisal exhibits a tendency towards

a weak negative correlation with RT (Figure 4), counter to our expectation.

3.2. Long vowel duration

As syllable surprisal values did not vary in low frequency words containing short vowels, this was problematic in fitting a model to the data to examine the interaction between WORD FREQUENCY and SYLLABLE SURPRISAL. Therefore, our acoustic analysis focused on the data set containing only long vowels. The final model had the structure of $\sim$ SYLLABLE SURPRISAL * WORD FREQUENCY * SYLLABLE * POSITION + (SYLLABLE + POSITION | SPEAKER) + (1 | WORD). The model yielded four main effects: SYLLABLE SURPRISAL ($F(1, 9.15) = 9.02, p = .015$), SYLLABLE ($F(1, 14.32) = 54.83, p < .001$), POSITION ($F(1, 37.88) = 124, p < .001$), with two 2-way interactions: SYLLABLE SURPRISAL-BY-SYLLABLE ($F(1, 9.15) = 33.07, p < .001$), and SYLLABLE-BY-POSITION ($F(1, 684.3) = 5.12, p = .024$), as well as one 3-way interaction: SYLLABLE SURPRISAL-BY-WORD FREQUENCY-BY-SYLLABLE ($F(1, 9.15) = 20.29, p = .001$). To examine the 3-way interaction, monosyllabic and disyllabic items were analyzed separately, like that carried out in section 3.1 for comparison and interpretation purposes.

3.2.1. Monosyllabic words

The optimal model fit had the structure of $\sim$ WORD FREQUENCY + SYLLABLE SURPRISAL + POSITION + (POSITION | SPEAKER) + (1 | WORD), because the model with interactions did not differ from that without in terms of the result from the likelihood ratio test. The optimal model yielded a significant effect of POSITION ($F(1, 20.29) = 79.59, p < .001$), with longer vowel duration in final than medial position. No other effects were observed.

3.2.2. Disyllabic words

The optimal model fit had the structure of $\sim$ WORD FREQUENCY * SYLLABLE SURPRISAL * POSITION + (POSITION + WORD FREQUENCY | SPEAKER) + (1 | WORD), based on the likelihood ratio test: $\chi^2(4) = 11.1, p = .026$. The results revealed significant main effects of WORD FREQUENCY ($F(1, 5.61) = 8.49, p = .03$), SYLLABLE SURPRISAL ($F(1, 5.61) = 29.5, p = .002$), and POSITION ($F(1, 24.02) = 174.7, p < .001$), with one 2-way interaction: SYLLABLE SURPRISAL-BY-WORD FREQUENCY ($F(1, 408.17) = 14.8, p = .009$). Figure 5 shows the interaction. Vowel duration tends to negatively correlate with syllable surprisal, with the effect being stronger in low frequency words, counter to our expectation that high syllable surprisal will trigger long acoustic vowel duration.

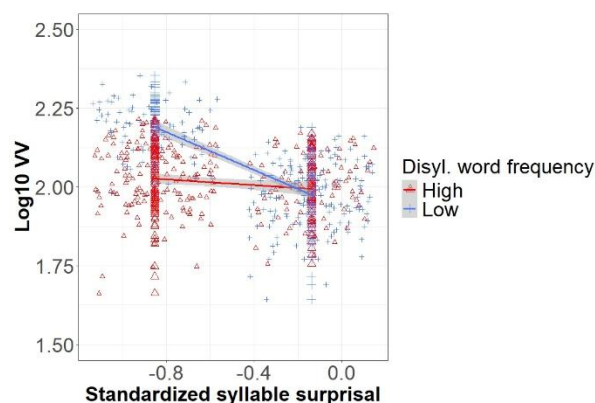


Figure 5: Long vowel duration (log) as a function of standardized syllable surprisal in high- and low-frequency disyllabic words.

4. Discussion

Overall, our RT and acoustic long vowel duration findings revealed interaction effects, rather than the hypothesized additive effects. As regards RT, the observed pattern is conditional on syllable type: monosyllabic or disyllabic words. For monosyllabic words, RT patterns follow syllable surprisal, perhaps reflecting syllable encoding; however, the direction of the syllable surprisal effect depends on whether the word has high frequency or low frequency. When it is a high frequency word, high syllable surprisal slows RT, suggesting that a speaker might need more time to resolve the conflict between a frequent word and an infrequent syllable. However, when it is a low frequency word, high syllable surprisal speeds up RT because there is no need to reconcile precompiled forms across word and syllable levels. For disyllabic words, high syllable surprisal also speeds up RT irrespective of word frequency. Our RT findings partially support the notion of a ‘stored’ syllable, but only for the high frequency monosyllabic words (in line with [9, 12]). It remains unclear why the syllable surprisal effect is sensitive to the lexical frequency of monosyllabic words, but not that of disyllabic words. Perhaps this is related to the relatively restricted range of syllable surprisal in LF monosyllabic words (see Figure 2). This limitation should be addressed in future experiments to extend the range of syllable surprisal.

As regards the acoustic long vowel duration measure, the effect of syllable surprisal is also conditional on word frequency. This interaction is only present for disyllabic words, not monosyllabic ones. Surprisingly, the direction of the syllable surprisal effect for low frequency disyllabic words is counter to our expectation, with shorter duration for high surprisal syllables. Alternatively, the vowel duration pattern can be interpreted as the effect of word frequency, conditional on syllable surprisal. That is, vowel duration is longer in low frequency than high frequency words, when syllable surprisal is low. However, the effect of word frequency disappears when syllable surprisal is high. Perhaps high syllable surprisal may have overpowered the effect of word frequency; whereas low syllable surprisal could not. In addition, the lack of a syllable surprisal effect for the high frequency disyllabic words might arise from the interplay of the intrinsic constraint of long vowels, polysyllabic shortening, word frequency and syllable surprisal. Polysyllabic shortening and high frequency word shortening might make it hard for the syllable surprisal effect to manifest without the cost of maintaining intrinsic vowel duration.

Although word frequency and syllable surprisal affect RT and acoustic duration, these effects are manifest in different syllable types. The effects of word frequency and syllable surprisal on RT appear in monosyllabic and disyllabic words respectively; whereas the interaction of word frequency and syllable surprisal appear on the acoustic long vowel duration of the disyllabic words only. These findings suggest that RT and the acoustic measure do not mirror one another and might reflect different aspects of the speech production planning process. RT might reflect how a unit and what kind of units need to be reached prior to speaking; whereas the acoustic measure characterizes the interplay of structural and usage constraints on phonetic specifications of the acoustic output.

5. Conclusions

In sum, the current study examined word frequency and syllable surprisal on speech initiation response time (RT) and acoustic vowel duration in the context of the psycholinguistic model of speech production in [14]. Counter to the hypothesized additive effects of word frequency and syllable surprisal, our findings, based on 2 different measurement metrics, generally reveal a cross-level interaction effect. These findings cast doubt on the assumption of ‘discrete, staged and serial’ processing among levels in [14], and are rather more in agreement with ‘interactive’ types of speech production accounts.

6. Acknowledgements

The research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 232722074 – SFB 1102 Information Density and Linguistic Encoding.

7. Generative AI Disclosure

No generative AI has been used.

8. References

- [1] M. Aylett, and A. Turk, (2004). “The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence duration in spontaneous speech,” *Language and Speech*, 47, pp. 31-56, 2004.
- [2] M. Aylett and A. Turk, “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei,” *Journal of the Acoustical Society of America*, 119 (5): 3048-3058, 2006.
- [3] I. Arnon, and U. Cohen Priva, “More than words: The effect of multi-word frequency and constituency on phonetic duration,” *Language and Speech*, 56 (3), pp. 349-371, 2013.
- [4] M. Pluymaekers, M. Ernestus, R. H. Baayen, “Effects of word frequency on the acoustic durations of affixes,” in *Proc. INTERSPEECH 2006 – ICSLP*.... Pp. 953-956.
- [5] S. Gahl, “Time and thyme are not homophones: word durations in spontaneous speech,” *Language*, 84, pp. 474-496, 2008.
- [6] A. H. Kawamoto, C. T. Kello, I. Higareda, and J. Q. Vu, “Parallel processing and initial phoneme criterion in naming words: evidence from frequency effects on onset and rime duration,” *Journal of Experimental Psychology*, 25, pp 362-381, 1999.
- [7] S. T. Piantadosi, H. Tily, E. Gibson, “Word lengths are optimized for efficient communication,” *Proc. Natl. Acad. Sci.* 108 (9), 3526–3529, 2011.
- [8] O. Ibrahim, I. Yuen, M. van Os, B. Andreeva, and B. Möbius, “The combined effects of contextual predictability and noise on the acoustic realization of German syllables,” *The Journal of the Acoustical Society of America*, 152, pp. 911-920, 2022.
- [9] J. Cholin, W. J. M. Levelt and N. O. Schiller, “Effects of syllable frequency in speech production,” *Cognition*, 99, pp. 205-235, 2006.
- [10] K. Croot, G. Lalas, B. Biedermann, K. Rastle, K. Jones, and J. Cholin, “Syllable frequency effects in immediate but not delayed syllable naming in English,” *Language, Cognition and Neuroscience*, 32:9, 1119-1132, 2006.
- [11] J. D. Jescheniak, and W. J. M. Levelt, “Word frequency effects in speech production: retrieval of syntactic information and of phonological form,” *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, pp. 824-843, 1994.
- [12] M. Laganaro, and F-X. Alario 2006. “On the locus of the syllable frequency effect in speech production,” *Journal of Memory and Language*, 55, 178-196.
- [13] P. Mousikou, and K. Rastle, “Lexical frequency effects on articulation: a comparison of picture naming and reading aloud,” *Frontiers in Psychology*, 6, 2015. <https://doi.org/10.3389/fpsyg.2015.01571>
- [14] W. J. M. Levelt, A. Roelofs, and A. S. Meyer, (1999): “A theory of lexical access in speech production,” *Behav Brain Sci* 22:1–38, 1999.
- [15] M. Brysbaert, SUBTLEX-DE. Retrieved from osf.io/py9ba
- [16] R. H. Baayen, R. Piepenbrock, and L. Gulikers, CELEX2 LDC96L14. Philadelphia: Linguistic Data Consortium, 1995.
- [17] M. Baroni, S. Bernardini, A. Ferraresi, and E. Zanchetta, “The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora,” *Language Resources and Evaluation*, 43(3), 209-226, 2009.
- [18] G. Möhler, A. Schweitzer, M. Breitenbücher, and M. Barbisch, IMS German Festival (version 1.2-os), University of Stuttgart: Institut für Maschinelle Sprachverarbeitung (IMS) retrieved 2020-01-02, 2000.
- [19] A. Stolcke, “SRILM – an extensible language modeling toolkit,” in *Proceedings of INTERSPEECH 2002*, vol. 2. Denver, Colorado, 2002, pp. 901-904.
- [20] I. H. Witten, and T. C. Bell, “The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression,” *IEEE Transactions on Information Theory*, 37(4), 1085-1094, 1991.
- [21] P. Boersma, and D. Weenink, Praat: doing phonetics by computer [Computer program].
- [22] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, 67(1), pp. 1-48, 2015.
- [23] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest Package: Tests in Linear Mixed Effects Models,” *Journal of Statistical Software*, 82(13), pp. 1–26, 2017.
- [24] R Core Team R: A Language and Environment for Statistical Computing, 2022. [Computer program].