CONSONANTS, VOWELS, AND LEXICAL REPRESENTATIONS: EVIDENCE FROM AUDITORY LEXICAL DECISION

Scott James Perry, Gabriela Holko, Matthew C. Kelley, and Benjamin V. Tucker
University of Alberta

1. Introduction

For a number of languages in the world, it has been asserted that consonants are more closely related to lexical representations than vowels, based on findings from thirteen languages across seven language families (for a review see Nazzi & Cutler, 2019). Several different experimental methodologies have been employed with adults to address this research question. The present study investigates the relationship between consonants, vowels, and lexical representations by analyzing how response latencies in auditory lexical decision data from both native and non-native listeners are influenced by the proportions of vowels and consonants within a word. Since lexical access is occurring during auditory lexical decision, we believe that it is an appropriate source of information for investigating this phenomenon.

A prominent experimental paradigm that has been used to argue that consonants are more closely tied to lexical representations than vowels is word reconstruction. Van Ooijen (1996) conducted the first such experiment in British English. In this paradigm, participants are presented with pseudowords that differ from real words by one segment, with the items designed in such a way that either a vowel or a consonant can be changed to form an existing word. Participants are asked to change one sound to make an existing word. For example, they could be presented with a recording of the pseudoword ⟨shevel⟩ [ʃɛvəl], from which they could change a consonant to arrive at the word *level*, or a vowel to arrive at the word *shovel* (van Ooijen, 1996). Participants chose to change vowels more often, and when asked to change consonants they made more errors and took longer to respond. Van Ooijen interpreted this pattern as indicating that vowels contribute less to the lexical identity of words than consonants. They argued that if acoustic input does not trigger lexical access, as in the case of pseudowords, then participants allow the vowel to be more "mutable" and accessing a word with a vowel substitution happens faster than accessing a word with a consonant substitution.

This word reconstruction paradigm has been used to investigate other languages, with consonants being argued to be more important for lexical representations in Dutch, Japanese, Spanish, and American English (Cutler et al., 2000; Marks et al., 2002; Cutler & Otake, 2002; Moates et al., 2002). The replication of this finding in various languages provided evidence against potential alternative explanations such as the composition of vowel/consonant inventories, phonological vowel reduction, and the presence/absence of lexical stress and lexical pitch accents.

New et al. (2008) postulated that this finding would not hold in languages with lexical tone, as lexical information is carried by the vowel. Wiener and Turnbull (2016)

Actes du congrès annuel de l'Association canadienne de linguistique 2021.
Proceedings of the 2021 annual conference of the Canadian Linguistic Association.

found evidence supporting this claim from a word reconstruction experiment in Mandarin. They found that participants were most likely to change the lexical tone instead of changing an entire consonant or vowel segment. According to Nazzi and Cutler (2019), this is the only such study at the time of writing to not observe this preference for changing vowels, although there were some differences in stimuli design that allow for alternative explanations for these findings.

Word reconstruction is not the only experimental paradigm that has been employed to investigate this line of research. Other experimental paradigms used to support the relationship between consonants and lexical representations have included artificial language learning experiments and segmenting real words from a continuous stream of speech. During artificial language learning experiments, participants are presented with a long string of repeating words and are asked to identify the words that make up the speech stream. The general finding of this research is that these invented words are more readily parsed from the speech stream when the transitions between the words are marked with consonants rather than vowels (e.g., Bonatti et al., 2005), even when the probability of the consonant and vowel materials were the same (Mehler et al., 2006).

When segmenting existing words out of larger contexts, listeners are presented with words embedded with additional phonetic content either before or after the word. Listeners more readily recognize words in the stream when there is additional vocalic content. For example, Norris et al. (1997) and Kearns et al. (2002) found that sleep is more readily recognized when embedded in the pseudoword *sleepnah* than in *sleept*.

While research into the relationship between segment type and lexical representations has focused almost entirely on the lexical processing of native speakers, Wiener (2020) ran word reconstruction tasks in Mandarin and English with native speakers and second language (L2) learners of both languages. In Mandarin, both native speakers and L2 learners were more likely to change a consonant, contrary to previous work in Mandarin by Wiener and Turnbull (2016). In English, both native and L2 listeners were more likely to change a vowel. Even L2 learners of intermediate levels of proficiency patterned with native speakers in the segments they chose to change, showing different strategies in their two languages.

The present study aims to investigate the relationship between consonants, vowels, and lexical representations by analyzing the response latencies in a large database of auditory lexical decision data for Canadian English. While priming experiments in visual lexical decision have been used to investigate this research topic previously (e.g., Cutler et al., 1999), to our knowledge, the response latencies in auditory lexical decision have not been. We believe this is a potential source of information that is relevant, and our linking hypothesis is as follows: if consonants are more closely related to lexical representations than vowels are, words with more consonants and less vowels should be responded to faster, and words with a higher ratio of vowels to consonants should be responded to more slowly. Our research questions are the following. First, does the proportion of vowels in a word affect the response latency in an auditory lexical decision task in native English listeners? Second, is the effect of the proportion of vowels the same for non-native listeners whose native language has lexical tone?

Based on the literature from different languages and various experiments, we predicted that native English speakers would respond to words with more consonants faster, and words with more vowels slower. For non-native listeners, we predicted that non-native speakers would present the same pattern of the effect as native speakers based on the findings of Wiener (2020).

2. Methods

The data for the present study comes from the Massive Auditory Lexical Decision database (MALD; Tucker et al., 2019; Perry et al., in prep). MALD consists of auditory lexical decision data of responses to 26,793 words and 9,592 pseudowords. The present dataset consists of a combination of the MALD 1.1 dataset (Tucker et al., 2019) and a to-be-released MALD L2 listener dataset (Perry et al., in prep). For the present analysis, participants were included if they gave English or Mandarin as their native language. In the present analysis, we only analyzed the response latencies. As is common practice in analyzing response latencies in lexical decision, only accurate responses to real words were analyzed. Response latencies less than 200ms from word onset were discarded as being too fast to be legitimate responses.

2.1 Participants

The participants included in the present study were those in the MALD 1.1 and L2 data who reported their native language to be either English or Mandarin (N=640). There were more native English speakers (n=495) than native Mandarin speakers (n=145). The ages of the two groups were similar, with English natives having an average age in years of 20.45 (SD=3.80), and Mandarin natives 20.68 (SD=2.96). English proficiency was also collected on a scale from one to five. English native speakers had an average of 4.99 (SD=0.04). Mandarin natives had lower proficiency and showed more variation, with an average of 3.25 (SD=0.77).

2.2 Calculating proportion of vowels

We calculated the proportion of consonants and vowels in each item for all real words in the MALD data. This was done in two ways. The first method, which we refer to as the "phone method", used the phonetic transcriptions included in the MALD data, calculating the total number of vowels in each item and dividing it by the total number of phones. The second method, which we refer to as the "acoustic method", took advantage of the available Praat TextGrids available for MALD items. These text grids provide phone-level segmentations for all items in the data. The acoustic method calculated the total length of all vowels in each word and divided that by the total duration. Two words were removed from the analysis as they did not have corresponding TextGrids.

The density distributions for both methods are plotted in Figure 1. The shape of the distribution of the phone method is due to the number of values in the calculation; there is a limited number of possible values for vowels and phones contained within a word. The acoustic method more closely approximates a normal distribution, as the temporal duration of the vowels present compared to the overall duration had much more variability.

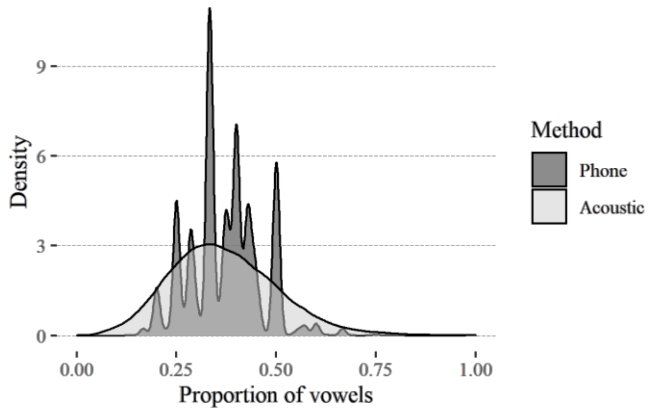

Figure 1. The density distributions of the proportion of vowels in the MALD data for all real words. The density distribution as calculated by the phonetic transcriptions is in dark gray, and the density distribution as calculated by the acoustic segmentation is in light gray.

2.3 Statistical modelling

To model the effect of the proportion of vowels on response latencies, we fit a series of generalized additive mixed-models using the mgcv (version 1.8-36, Wood, 2012) package in R (version 4.1.1, R Core Team, 2021). We first fit simple models with only the proportion of vowels, calculated by the acoustic method, as the only smooth term, with random intercepts for participants and a random slope for the proportion of vowels by participant. This was done to investigate potential suppression effects. We did not fit random intercepts by item due to the large number of items. After fitting the simple model, a full model was fit that included a number of covariates known to influence response latencies in auditory lexical decision. Three of these predictors were lexical in nature: word frequency (calculated from the SUBTLEX corpus), duration of the word in milliseconds, and phonological neighborhood density. The remaining covariates were experimental in nature: word run length (the number of previous items in a row to which that participant had responded that it was a word) and trial number, which gives information about how participants change over the time-course of the experiment. All lexical covariates were log-transformed and scaled in all models following standard practice. All experimental variables were scaled. All covariates were modeled as smooth terms. The full models with covariates were also fit using the proportion of vowels as calculated by the phone method.

These models were first fit to a subset of the MALD data that only included native speakers of English. The next step included native speakers of Mandarin in the models. In this case, separate smooth terms were fitted to specify the interaction between all lexical-level predictors and participants' native language. The analysis materials, including the R script and the data-set of the acoustic information, are available as supplementary materials here: https://doi.org/10.7939/r3-8sey-d843.

3. Results

The results of the statistical modelling are reported in the following sections. These models all report the results of modelling the proportion of vowels using the acoustic method, as this method contains more information about the acoustic signal presented to participants. The models fit using this variable as calculated with the phone method are available in the supplementary materials but not reported here in the interests of space. The model predictions across the two methods were similar in their interpretations. Model summaries are located in the Appendix.

3.1 Native-only model results

For native speakers, the smooth term for the proportion of vowels in the word was statistically significant in the model without additional covariates (edf=7.092, p=<0.001) and in the model with additional covariates (edf=4.997, p=0.007). However, the effect of this predictor changes with the addition of covariates, which can be seen in Figure 2. When modelled alone, words with average proportions of vowels and consonants are responded to more slowly than words with more or fewer vowels than average. In the model with covariates, we see that within two standard deviations of the average proportion of vowels there is no difference, but words with a higher proportion of vowels are responded to more slowly.

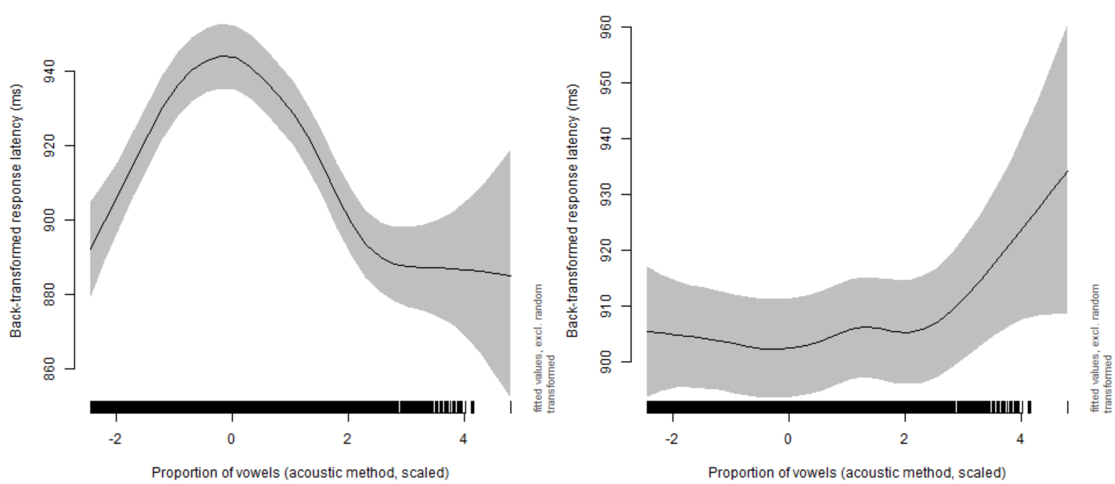

Figure 2. Left: The predicted effect for the proportion of vowels on the response latencies of native English speakers in a model with no covariates. Right: The predicted effect for the proportion of vowels in a model with additional control covariates. In both plots y-axis is back-transformed response latency, x-axis is scaled proportion of vowels as calculated

by the acoustic method. The rug at the bottom of the plot indicates where data points were located.

3.2 Model with non-native listeners

For L2 listeners, the effect of the proportion of vowels was not significant in the model with covariates (edf=1.00, p=0.285). The effect for both native and non-native listeners from the model containing both is plotted in Figure 3. As L2 learners were slower overall, they were significantly different from native speakers for most values of the proportion of vowels. Only in words with high proportions of vowels was the difference between the two groups not statistically significant, as can be seen in the difference plot on the right side of Figure 3, which subtracts the L2 curve from the native curve.

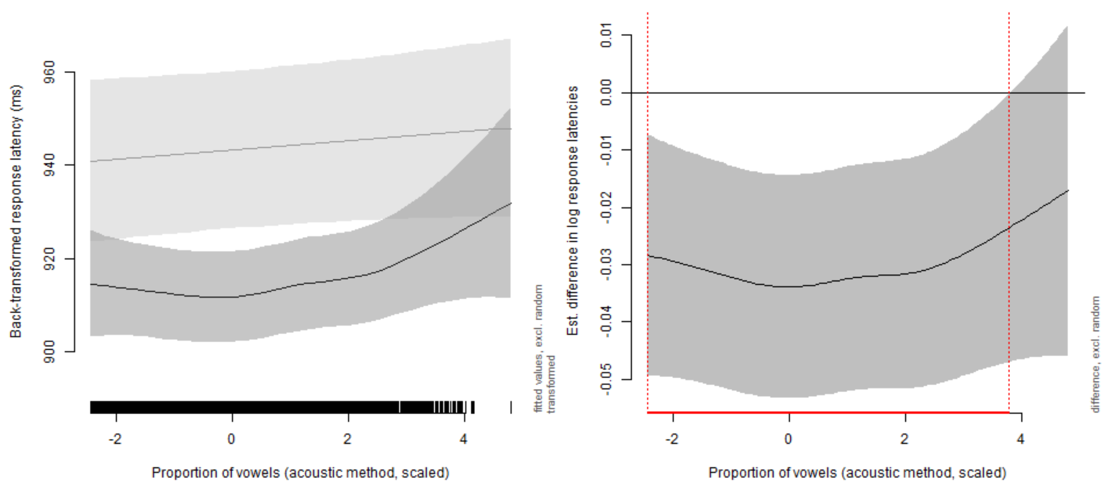

Figure 3. Left: The predicted effect for the proportion of vowels on the response latencies of native English speakers (dark gray) and native Mandarin listeners (light gray). Y-axis is back-transformed response latency, x-axis is scaled proportion of vowels as calculated by the acoustic method. Rug at bottom of plot indicates where data was located. Right: The difference plot between the native English and native Mandarin listeners. The section outlined in red indicates where the difference between the two curves presented on the left is statistically significant.

The results of the modelling using the phone method provided similar results to those reported here. The effect for natives followed the same nonlinear pattern with a similar interpretation, the only difference being that the slight increase in reaction times for native speakers in words with a low proportion of vowels is more pronounced. The effect for L2 listeners was similarly non-significant. These models are not reported in full here for reasons of space, but can be found at the end of the analysis script included in the supplementary materials.

4. Discussion

The present study aimed to bring a new type of experimental data to bear on the relationships that consonants and vowels have to lexical representations in native and non-native speakers. Response latencies from a database of English auditory lexical decision data were analyzed for native English listeners and native Mandarin/L2 English listeners. Our first prediction was as follows: if consonants are more closely related to lexical representations, then words with proportionally more consonants should be recognized and responded to more quickly, and words with proportionally more vowels should be responded to more slowly.

In our first analysis, we focused on this effect in native listeners of English only. We found a significant effect of the proportion of vowels within a word. The effect was non-linear, with the proportion of vowels not affecting response latencies within two standard deviations of the average proportions of vowels for English. It was only words with a very high proportion of vowels that were responded to more slowly.

Our second prediction related to differences between native and non-native listeners and predicted that L2 listeners would pattern similarly to native speakers, as previous research has claimed that L2 learners follow non-native processing patterns. In our second analysis, we included L2 English listeners who gave Mandarin as their native language. There was no significant effect of the proportion of vowels on response latencies for these listeners, contrary to our predictions.

The results of the effect of the proportion of vowels in native speakers are not incompatible with the idea that consonants are more related to lexical representations. Our prediction based on this claim was that words with a higher proportion of vowels would be responded to more slowly. However, we also predicted that words with a higher proportion of consonants would be responded to more quickly, which the data does not support. It is important to note here that our hypothesis was phrased in terms of a linear effect but tested in a model that allowed for non-linear effects.

For both native and non-native listeners, it is important to point out that the effect, or lack thereof, of the proportion of vowels is dependent on the other variables included in the model. When only the proportion of vowels was included as a predictor, the effect was much different, with the slowest response latencies being found at average proportions of consonants and vowels, and responses being faster for higher proportions of both consonants and vowels in a quasi-quadratic shape. This effect changed because the proportion of vowels was related significantly to all other lexical predictors that served as covariates. Which variables should or should not be included in a statistical model to find the true effect of a predictor of interest is a non-trivial issue. In this case, after the inclusion of duration and frequency, the effect remained relatively stable for native listeners or disappeared for L2 listeners.

The effect of the proportion of vowels only showed an effect in words that were not typical in the proportion of vowels to consonants. As we had a massive sample of English words in this study, it is likely that the properties of our stimuli closely reflect the distribution of the English language more broadly. This could mean that if a much higher than average proportion of consonants does speed up responses, this effect would have to be investigated in a language that presents a different distribution of these proportions. While these findings are more nuanced than previous claims would have allowed us to predict, they may support the idea that consonants are more closely related to lexical representations in English, albeit with an added caveat that words with relatively typical proportions of vowels and consonants are all responded to similarly.

The observed effect is compatible with the phenomenon we intended to investigate, but alternative explanations are of course possible. Words with much higher proportions of vowels in English could be thought of as less wordlike, in that they are unlike the majority of English words in their composition. If what we are observing in this study is a wordlikeness effect (Bailey & Hahn, 2001), we would expect findings for other languages to have a U-shaped effect, with words that have many more consonants than average to have slower response latencies. If we are observing the link between consonants, vowels, and lexical representations as intended, we would expect words with many more consonants than average to be responded to more quickly.

The difference between the findings for native and non-native listeners was contrary to our predictions, as previous research demonstrated that L2 learners of even intermediate levels of proficiency behaved similarly to native speakers in a word reconstruction task in both English and Mandarin. We must be careful of interpreting the absence of evidence as evidence of absence, but it is possible that there is no such effect for native speakers. This could be a matter of a task effect, as similar performance in one task does not necessarily mean that results will be similar in another. Before exploring a unifying explanation of both phenomena, replication of both studies would be important.

Perhaps the most logical next step in this research is to replicate this finding with other databases of lexical decision data in English, such as Auditory English Lexicon Project (Goh et al., 2020). Another potential avenue is to test this effect in other languages that are purported to also favour consonants in lexical representations, such as Dutch (with BALDEY; Ernestus & Cutler, 2015), Spanish, or Japanese. It would be useful to test this hypothesis in languages where the average proportion of vowels is different from English. It would also be important to use non-linear regression methods in this case as well, as using linear regression in our case would have confirmed both of our predictions when only one of them was supported when effects were allowed to depart from linear trends.

Another potential follow-up would be to investigate how this effect changes in L2 learners as a function of proficiency or quantity or quality of the input. While proficiency is present in the MALD database, this data was not included in the present analysis as treating this ordinal predictor as continuous or categorical is not statistically appropriate (Bürkner & Charpentier, 2020), and the size of the data-set made implementing appropriate methods difficult. It may be that the L2 listeners in our sample had not received a sufficient amount of input to approximate native-like processing of English words, or that looking more in-depth into other factors such as the age of onset of acquisition or language use could have shown different processing patterns in different types of bilinguals.

References

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods?. *Journal of Memory and Language, 44*(4), 568-591.

Bonatti, L. L., Pena, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*(6), 451-459.

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology, 73*(3), 420–451. https://doi.org/10.1111/bmsp.12195

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., & Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & cognition, 28*(5), 746-755.

Cutler, A., Van Ooijen, B., & Norris, D. (1999). Vowels, consonants, and lexical activation. In The Fourteenth International Congress of Phonetic Sciences (pp. 2053-2056). University of California.

Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of memory and language, 46*(2), 296-322.

Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *The Quarterly Journal of Experimental Psychology*, *68*(8), 1469–1488.

Goh, W. D., Yap, M. J., & Chee, Q. W. (2020). The Auditory English Lexicon Project: A multi-talker, multi-region psycholinguistic database of 10,170 spoken words and nonwords. *Behavior Research Methods*.

Marks, E. A., Moates, D. R., Bond, Z. S., & Stockmal, V. (2002). Word reconstruction and consonant features in English and Spanish. *Linguistics, 40*(2).

Moates, D. R., Bond, Z. S., & Stockmal, V. (2002). Phoneme frequency in spoken word reconstruction. *Laboratory Phonology 7*, 141.

Nazzi, T., & Cutler, A. (2019). How Consonants and Vowels Shape Spoken-Language Recognition. *Annual Review of Linguistics, 5*(1), 25–47. https://doi.org/10.1146/annurev-linguistics-011718-011919

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio, 2*(2), 203-230.

New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science, 19*(12), 1223-1227.

Perry, S.J., Nijveld, A., & Tucker, B.V. (in prep.). *Multilingual spoken word recognition: A megastudy approach*. Manuscript in preparation.

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2019). The massive auditory lexical decision (MALD) database. *Behavior research methods, 51*(3), 1187-1204.

van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition, 24*(5), 573-583.

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and speech, 59*(1), 59-82.

Wiener, S. (2020). Second language learners develop non-native lexical processing biases. *Bilingualism: Language & Cognition, 23*(1), 119-130.

Wood, S. (2012). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. http://cran.r-project.org/web/packages/mgcv/index.html

Appendix

Table A1. Regression summary table for the model containing only native speakers and no additional covariates. Parametric terms are listed first, followed by smooth terms. Numbers have been rounded to three decimal places.

Parametric terms	Estimate	SE	*t*-value	*p*-value
(Intercept)	6.839	0.004	1485	<0.001
Smooth terms	edf	ref.df	F	*p*-value
s(Proportion of Vowels)	7.092	8.176	93.81	<0.001
s(Subject)	488.061	495	84.53	<0.001
s(Subject,Proportion of Vowels)	0.001	495	0.00	0.732

Table A2. Regression summary table for the model containing only native speakers with additional control covariates. Parametric terms are listed first, followed by smooth terms. Numbers have been rounded to three decimal places.

Parametric terms	Estimate	SE	*t*-value	*p*-value
(Intercept)	6.837	0.005	1479	<0.001
Smooth terms	edf	ref.df	F	*p*-value
s(Proportion of Vowels)	4.997	6.171	3.009	<0.007
s(Word Frequency)	6.061	7.026	944.768	<0.001

s(Word Duration)	7.260	8.187	1297.468	<0.001
s(Phonological ND)	6.679	7.493	166.753	<0.001
s(Word Run Length)	4.107	4.928	145.922	<0.001
s(Trial)	8.732	8.979	288.229	<0.001
s(Subject)	488.940	495	96.881	<0.001
s(Subject,Proportion of Vowels)	1.715	495	0.003	0.463

Table A3. Regression summary table for the model containing native and non-native speakers and no additional covariates. Parametric terms are listed first, followed by smooth terms. Native English speakers have been absorbed into the intercept in the parametric terms. Numbers have been rounded to three decimal places.

Parametric terms	Estimate	SE	*t*-value	*p*-value
(Intercept)	6.839	0.004	1485	<0.001
NativeLangMandarin	0.018	0.009	1.891	0.059
Smooth terms	edf	ref.df	F	*p*-value
s(Proportion of Vowels):NativeLangEnglish	6.939	8.056	85.72	<0.001
s(Proportion of Vowels):NativeLangMandarin	6.362	7.564	42.93	<0.001
s(Subject)	630.134	639	91.35	<0.001
s(Subject,Proportion of Vowels)	0.005	639	0.00	0.871

Table A4. Regression summary table for the model containing native and non-native speakers with additional control covariates. Parametric terms are listed first, followed by smooth terms. Native English speakers have been absorbed into the intercept in the parametric terms. Numbers have been rounded to three decimal places.

Parametric terms	Estimate	SE	*t*-value	*p*-value
(Intercept)	6.837	0.005	1479	<0.001
NativeLangMandarin	0.014	0.009	1.488	0.137
Smooth terms	edf	ref.df	F	*p*-value
s(Proportion of Vowels)	4.997	6.171	3.009	<0.007
s(Word Frequency)	6.061	7.026	944.768	<0.001
s(Word Duration)	7.260	8.187	1297.468	<0.001
s(Phonological ND)	6.679	7.493	166.753	<0.001
s(Word Run Length)	4.107	4.928	145.922	<0.001
s(Trial)	8.732	8.979	288.229	<0.001
s(Subject)	488.940	495	96.881	<0.001
s(Subject,Proportion of Vowels)	1.715	495	0.003	0.463