

RESEARCH ARTICLE

The recognition of spoken pseudowords

Matthew C. Kelley^{a,b} and Benjamin V. Tucker^a

^aDepartment of Linguistics, University of Alberta, Edmonton, AB T6G 2E7, Canada

^bDepartment of Linguistics, University of Washington, Seattle, WA 98195-0005, USA

ARTICLE HISTORY

Compiled March 10, 2022

ABSTRACT

Pseudowords are used as stimuli in many psycholinguistic experiments. Yet, pseudoword responses remain largely under-researched. To better understand the cognitive processing of pseudowords, we analysed the pseudoword responses in the Massive Auditory Lexical Decision megastudy data set (Tucker et al., 2019). Linguistic characteristics that influence the processing of real English words—namely, phonotactic probability, phonological neighbourhood density, uniqueness point, and morphological complexity—were also found to influence the processing time of spoken pseudowords. Subsequently, we analysed how the linguistic characteristics of non-unique portions of pseudowords influence processing time. We again found that the named linguistic characteristics affect processing time, highlighting the dynamicity of activation and competition. We argue these findings also speak to learning new words and spoken word recognition generally. We then discuss what aspects of pseudoword recognition a full model of spoken word recognition must account for. We finish with a re-description of the auditory lexical decision task in light of our results.

KEYWORDS

Phonetics; psycholinguistics; auditory lexical decision; spoken word recognition; pseudowords; mental lexicon

CONTACT Matthew C. Kelley. Email: mattck@uw.edu

1. Introduction

Certain linguistic characteristics of a word, such as its lexical frequency, can have strong effects on how the word is processed. For example, lexical frequency significantly affects the speed with which a word is recognised (Dahan et al., 2001; Dupoux and Mehler, 1990; Ernestus and Cutler, 2015; Glanzer and Bowles, 1976; Howes, 1954). When carefully interpreted, linguistic characteristics like lexical frequency can speak to the structure and properties of lexical representations. Returning to lexical frequency, its effect has been interpreted in many ways, such as each word having a different resting level of activation or each word having different connection strengths related to its frequency (Dahan et al., 2001). The present study continues the theme of investigating linguistic characteristics that affect the processing of speech signals to learn more about how spoken word recognition and lexical representation work. In the present study, we focus on the processing of phonotactically legal pseudowords, which are commonly used as distractors in experimental tasks designed for spoken word recognition research.

Here, we define a “pseudoword” as a phonotactically licit phone sequence that could form a word in a language but happens not to (such as *blick* in English), while “non-word” is used to mean a phone sequence that could not form a word in a language (such as *bnick* in English). Many spoken word recognition studies have used pseudowords or nonwords largely as distractors during experimental tasks such as lexical decision. In this way, responses to the real word stimuli involve linguistic processing. However, these studies tend to leave the responses to the pseudowords unanalysed. The few studies that have analysed pseudowords have investigated aspects of real word recognition. This focus is in spite of the fact that recognising pseudowords makes up a large portion—often half—of the behaviour of a participant in tasks like lexical decision. Furthermore, the studies that examine pseudowords use only a small number of word and pseudoword/nonword stimuli (e.g., Goldinger et al., 1989; Vitevitch and Luce, 1998). Overall, the restricted amount of research on pseudoword recognition points to a lack of knowledge about how the human speech processing system determines that an item is not in the lexicon. If the goal for models of spoken word recognition is to

account for the experimental tasks used in research (let alone spoken word recognition in non-laboratory environments), it is necessary to have an understanding of the full process of word and pseudoword recognition.

In the context of examining how speech signals are recognised, it may seem inconsequential or ecologically invalid to investigate the processing of pseudowords. However, an understanding of pseudoword recognition can be extended to how novel words are recognised and processed. For example, consider the case when an adult or child hears a word they've never encountered before in their native language. At that point, the new word is effectively a pseudoword (although the listener likely assumes that what they heard is a word and has meaning). Additionally, knowledge of pseudoword recognition can be used to model how speech processing mechanisms detect that an error has occurred when processing the speech signal. After all, an error when processing the signal could first cause the listener to hear a pseudoword before correcting. In general, understanding pseudoword recognition can account for various speech-related processes that occur when the word recognition system determines that what has been heard is not a word. Indeed, upon hearing an ostensible pseudoword in conversation, a listener needs to determine if the incoming signal is a misperception of a word or if it is a word they have not previously encountered.

In experiments like lexical decision where pseudoword recognition occurs, a number of processes are at play. Principally, there is a recognition process whereby a listener discerns the identity of the auditory stimulus. At some level, the listener determines what the wordform is, whether that involves string recognition or some whole-form recognition, all while lexical competition is ongoing. Subsequently, a listener renders a judgement about whether the item is a real word or not. In contrast, in a typical conversation, a listener does not hear isolated words and is continuously integrating the auditory input into their representation of the discourse. Furthermore, listeners generally assume everything they hear is meaningful. Although, listeners must also have some capacity to detect that something they heard is not in their mental lexicon based on their ability to recover from a perception error that results in what is effectively a pseudoword. Naturally, then, care is needed when inferences are made about cognitive processes on the basis of a task like lexical decision and many other

psycholinguistic tasks. The judgement process is distinct from recognition but may also affect the listener's responses. And, it must be noted that participating in an experiment is different from being in a conversation. However, lexical competition must also occur when hearing at least the non-unique portions of pseudowords because there would be no way for a listener to know that they are hearing a pseudoword at that point.

Analysing participant responses to pseudowords in experimental settings has a number of useful properties. The first among these is that it allows researchers to investigate the effects that properties of lexical items have on speech processing mechanisms in a more controlled manner. For example, since a good pseudoword stimulus should theoretically have a lexical frequency of 0, there should not be a confound of lexical frequency when investigating the effects of other characteristics of linguistic and/or phonetic sequences. Though, compare results from Hendrix and Sun (2020), who calculated a form of lexical frequency for visual, orthographic pseudowords through the number of results reported by Google searches for the pseudoword, suggesting that many pseudowords may appear at some point in the written language but are not recorded in a dictionary. However, from Google search results, it is difficult to tell whether the results for a given search string contain writing errors, are in a foreign language, or are proper nouns for which its lexical status for an entire language community is more questionable (e.g., *blick* <<https://web.archive.org/web/20200629162828/https://www.dickblick.com/>>). Additionally, Google attempts to determine user intent and return results based on that intent (such as with spelling correction and matching synonyms) rather than searching literally. As such, more investigation on the method of calculating the frequency of pseudowords using search engine results is needed. What's more, in the auditory modality, it is far more difficult to search for what may be instances of a pseudoword, which may be recognised as an error by the listener and corrected during the word recognition process. It still stands to reason, then, that lexical frequency is unlikely to be a confound when investigating different linguistic characteristics of experimental stimuli with pseudowords.

Pseudowords can also be designed to exhibit extreme values of lexical characteristics, as in Goldinger et al. (1989), Luce and Pisoni (1998), and Vitevitch and Luce (1998).

Note, though, that these studies did not investigate pseudoword processing in great detail, but instead used pseudowords to inform an understanding of strictly real word processing. Selecting such stimuli can prove more difficult with real words. As such, knowing if pseudowords are processed in similar ways as real words would serve as the basis for future experiments that use pseudowords as highly controlled stimuli in experiments investigating the effects of lexical characteristics on participant behaviour. This is especially true given results that pseudowords show variable wordlikeness, which itself is correlated with other linguistic characteristics (Bailey and Hahn, 2001).

To account for participant behaviour during linguistic experiments and what it may reveal about other aspects of language use, the processing of pseudowords *per se* merits investigation. There have been several recent studies on pseudoword processing. Yap et al. (2015) performed an extensive analysis of pseudowords in the visual modality, though linguistic predictors do not always have the same effect when compared across visual and auditory modalities (compare Luce and Pisoni, 1998; Yates et al., 2004). Chuang et al. (2019) and Chuang et al. (2020) modelled semantic activation that occurs while hearing pseudowords. These two studies couched their analyses in the linear discriminative learning and discriminative lexicon framework (Baayen et al., 2019). Additionally, Hendrix and Sun (2020) examined the process of recognising visual pseudowords using piecewise exponential additive mixed models; however, the auditory modality was not analysed in that study. In addition, while wordlikeness studies like Bailey and Hahn (2001) often use pseudowords as stimuli, they generally do not discuss the actual psychological process of recognising a pseudoword. Janse and Newman (2013) investigated the identification of auditory pseudowords through a transcription task. The selected stimuli were monosyllabic words, so multisyllabic words still need more attention. And, Vitevitch et al. (1997) looked at the effects of phonotactic probability on processing spoken nonwords, though only looking at bisyllabic words.

Because there is not an overwhelming amount of work yet on how pseudowords themselves are processed, the most fitting place to start is to work with what is known about processing real words. In that vein, four linguistic characteristics stand out as common predictors of behaviour in linguistic experiments that have implica-

tions for modelling human behaviour, especially in an auditory lexical decision task. These characteristics are phonotactic probability, phonological neighbourhood density, uniqueness point, and morphological complexity. These four characteristics were selected because they have been frequently used in previous research on spoken word recognition and pseudowords specifically, in the hope that our results will be maximally comparable to that body of research. These variables are used as proxies for different aspects of competition that have been shown to occur in spoken word recognition. Descriptions of these characteristics are elaborated in subsequent subsections.

1.1. Phonotactic probability

For real words, phonotactic probability is how likely a particular sequence of phones is, given a language’s phonotactic distribution. As a simplified example, using a combination of the CMU Pronouncing dictionary (Weide, 2005) and the Corpus of Contemporary American English (COCA) (Davies, 2008), it can be determined that /b/ is more frequent than /ŋ/ in English. As such, /bɪb/ is more probable than /bɪŋ/, when using single phones as the unit of analysis. Although, it is possible to use other units like diphones. The same principles apply to pseudowords. A pseudoword’s phonotactic probability could be an indication of how word-like it is, or otherwise, how well it fits a language’s phonotactic distribution.

Vitevitch et al. (1997) and Vitevitch and Luce (1998) have argued that human speech processing capabilities are aware of the frequency distributions of segment combinations. Overall, the trend for real words was that as the probability increased, the reaction time got slower, while the inverse was seen for the pseudowords, where the low-probability stimuli were responded to more slowly. However, they also found that the effect of phonotactic probability was overshadowed by a stimulus’s lexicality. Bailey and Hahn (2001) examined the contributions of both phonotactic probability and phonological neighbourhood density (discussed in the next section) to wordlikeness ratings of real words and pseudowords. They found that both phonotactic probability and phonological neighbourhood density had significant effects on participants’ responses. There is corroborating evidence from Frisch et al. (2000) that phonotactic probability influences the spoken word recognition process. It thus seems likely that

phonotactic probability has an effect on pseudoword recognition. And indeed, Janse and Newman (2013), Chuang et al. (2019), and Chuang et al. (2020) have shown that phonotactic probability influences response times to pseudo- and nonwords in auditory lexical decision tasks. However, Chuang et al. (2019) and Chuang et al. (2020) focus more on morphology and semantics than pseudoword recognition *per se*. More work needs to be done on how phonotactic probability affects processing and recognition, especially with multisyllabic pseudowords.

1.2. Phonological neighbourhood density

Phonological neighbourhood density is the number of phonological neighbours a word has. And, a phonological neighbour is typically defined as a lexical item that differs from the item in question by exactly 1 phoneme, as determined by Levenshtein distance (Luce, 1986; Luce and Pisoni, 1998). The phonological neighbourhood density for a given item, then, is a count of the phonological neighbours the item has. Generally, items with higher phonological neighbourhood density are expected to take longer to process because they will have more plausible competitors and, thus, take more time to process. Such effects have wide experimental support (Goldinger et al., 1989; Luce, 1986; Luce and Pisoni, 1998; Luce et al., 1990). See Vitevitch and Luce (2016) for a review. Note, however, that the neighbourhood density effect is not consistent cross-linguistically. Vitevitch and Rodríguez (2005) reported the opposite trend for spoken word recognition in Spanish.

While many spoken word recognition studies have reported effects of phonological neighbourhood density for pseudowords, the analyses were focused on understanding real word processing. To our knowledge, Janse (2009) is one of only a few studies to investigate the effect of phonological neighbourhood density on pseudoword processing itself. It was found that dense neighbourhoods increase response latencies to non-words in auditory lexical decision tasks in aphasic listeners, though no other structure was tested beyond CVC items. Chuang et al. (2019) and Chuang et al. (2020) also used phonological neighbourhood density as a predictor in their models when analysing multi-syllabic pseudowords, though not much explanation is given of its effect. Janse and Newman (2013) also used phonological neighbourhood density in their analyses.

Overall, though, the task of characterising the general effect of phonological neighbourhood density in pseudoword recognition remains open.

1.3. Uniqueness point

The segmental or phonological uniqueness point for a word is the segment after which only one item in the lexicon matches the stimulus being heard (Marslen-Wilson and Tyler, 1980; Marslen-Wilson and Welsh, 1978). Although, it has also been defined as when a lemma can be uniquely identified from all other possible outcomes (Ernestus and Cutler, 2015). Marslen-Wilson and Zwitserlood (1989) used uniqueness point in modelling responses in an audio-primed visual lexical decision task (cross-modal priming). Participants were primed with an auditory stimulus and then performed a visual lexical decision. Of the variety of experimental conditions that were assessed, a priming effect was only observed when the auditory stimulus was semantically related to the visual probe, and not, for example, if the auditory stimulus rhymed with the word related to the visual probe. They claim the speech processing mechanisms can use mismatches as they are heard to rule out possible candidates for recognising the word contained in the audio signal being heard, thus preventing the visual rhyming words from being primed.

Balling and Baayen (2012) found that the segmental uniqueness point has a significant effect and has an effect size comparable to lexical frequency in statistical models of responses in auditory lexical decision tasks. They also observed that a word's uniqueness point represents a moment of high entropy or surprisal, likely due to it being the moment at which the word is identifiable; after the uniqueness point, the information added by further segments in the word is redundant. The likely analogue for uniqueness point in a pseudoword is the point at which it can be determined that what the listener is hearing does not match any word in the lexicon; at that point, hearing further segments is redundant as well. The uniqueness point thus seems relevant to pseudoword processing, though there seems to have been less work on its effects for pseudowords than for some other linguistic characteristics. Note, though, that the uniqueness point represents the moment when there are 0 candidate words left for pseudowords, which may result in different processing effects for pseudowords. For this reason, it may not

be a moment of high surprisal as Balling and Baayen described for real words.

1.4. *Morphological complexity*

A variety of studies have found the morphological structure of words to affect language processing. Some evidence has come from priming experiments suggesting that words are obligatorily morphologically decomposed (Beyersmann et al., 2016; Lázaro et al., 2016; Rastle et al., 2004). In this process according to this account, all words are obligatorily parsed into morphemes during the recognition process. The parsing is blind as to whether or not a morphological parse is valid, such as breaking the English *corner* into *corn* and *-er*. Other studies have found that priming effects of obligatory decomposition can be explained in large part through other variables like semantic similarity (Feldman et al., 2009; Järvikivi and Pykkönen, 2011; Lõo and Järvikivi, 2019). There are fewer studies that have examined the effects of morphological complexity on pseudowords, though. Morris et al. (2011) found evidence that morphologically complex pseudowords like *flexify* from *flex* + *-ify* significantly primed participants' responses to target words in a visual lexical decision task. The evidence is still sparse, though Morris et al.'s findings suggest that pseudoword processing may be influenced by morphological complexity, as real words are.

Most studies investigating the effects of morphological structure on the processing of language use visual tasks, and not auditory tasks. While previous research suggests what trends may be observed in auditory processing, effects in the auditory domain still need to be examined. In part, this is because there exist perception differences between auditory and visual perception (Tucker et al., 2019). There are also differences in the stimuli where words that could be pseudo-complex orthographically would not be from a phonetic perspective. As an example, *brother* could be decomposed into *broth* + *-er* orthographically. This decomposition is implausible in the auditory domain because the first syllable of *brother* is [bɹʌð̃], while *broth* is [bɹʌθ]. In addition, Emmorey (1989) reported priming effects for spoken words that were morphologically related but semantically unrelated, but processing costs were not discussed.

Demonstrably, the effects of morphological complexity on auditory pseudoword processing need to be investigated. There are a handful of studies that have looked

at the relationship between morphological complexity and pseudoword processing. Moscoso del Prado Marti n et al. (2004) reported results in the visual domain that morphological complexity has an effect on pseudoword processing. In addition, Chuang et al. (2020) and Chuang et al. (2019) have found in the auditory domain that morphological complexity has an effect on pseudoword processing for a limited scope of morphological complexity measures. Overall, the preponderance of evidence from the visual and auditory domains suggest that the morphological complexity of real words affects language processing generally, whether through morphology itself or through other means such as semantic similarity. However, further investigation in the auditory domain is warranted with broader measures of morphological complexity.

1.5. The present study

In the present study, we seek to describe how pseudowords are recognised using the Massive Auditory Lexical Decision (MALD) data set (Tucker et al., 2019). Many of the studies discussed so far focused on learning more about real word processing and not pseudoword processing *per se* (Bailey and Hahn, 2001; Janse, 2009; Luce, 1986; Luce and Pisoni, 1998; Morris et al., 2011; Vitevitch and Luce, 1998; Vitevitch et al., 1997). That is, they were not primarily investigating the process of what goes on when listeners hear pseudowords (which is often half of what happens in an experiment involving a mix of pseudowords and real words). As well, the discussed studies that have looked at pseudoword processing have focused more on morphological and semantic investigations (Chuang et al., 2019, 2020) or have been in the visual modality (Hendrix and Sun, 2020; Moscoso del Prado Marti n et al., 2004). Understanding more about auditory pseudoword processing will relate to a number of linguistic phenomena, such as the aforementioned phenomena of learning a word a listener hasn’t heard before and recovering from perception errors. We note, here, that we don’t necessarily believe that there is a distinction between word and pseudoword processing. In both cases, lexical processing is occurring. But, we recognise that it is convenient from an explanatory perspective to separate words and pseudowords.

The overall motivating hypothesis is that pseudowords are processed using the same architecture as real words. This hypothesis breaks down into specific sub-hypotheses

related to the lexical characteristics discussed previously. For phonotactic probability, we expect that more phonotactically probable pseudowords will take longer to reject because they resemble common phonotactic tendencies in real words. For phonological neighbourhood density, we expect that pseudowords with more neighbours will take longer to reject because there are more competitors to narrow down from. For uniqueness point, we expect that pseudowords with later uniqueness points will take longer to respond to because more of the word will need to be heard before the rejection can be made. Finally, because some of the pseudowords being modelled are morphologically complex, we believe that pseudowords with more morphological complexity will take longer for listeners to reject. We use response latency in an auditory lexical decision task as a measure of processing time, with the linking hypothesis that longer response latencies relate to longer processing times due to a greater number of cognitive operations occurring.

The remainder of the paper is structured as follows. In the first of two analyses, more information about the data set being used is discussed. Then the modelling procedure for fitting a mixed-effects regression to the data is given. The first analysis then focuses on modelling the response time data, and the results are presented and discussed. However, standard analyses of response time data use linguistic characteristics calculated using the entire word or pseudoword. In effect, such analyses are only modelling endpoint characteristics of the stimuli, making the assumption that these endpoint characteristics are relevant for the entire process of recognising a word or pseudoword. In turn, this assumption leaves wide open the question of how the aspects of recognition that these linguistic characteristics represent evolve over time, such as the number of competitors that might be in competition only a few phones into the stimulus. This question is important to consider, especially because of the prominence of the activation and competition metaphor in spoken word recognition research.

Activation and competition are usually thought of as a dynamic process taking place over the entire time course and not all at once after the offset of a (pseudo)word. Indeed, there is a long history of evidence from gating studies (Grosjean, 1980) and visual world paradigm studies (Allopenna et al., 1998; Teruya and Kapatsinski, 2019)

that this process is dynamic and happens over time. As such, we believe it is crucial to understand how linguistic characteristics of a phonetic signal affect the dynamics of activation and competition over time—not just at stimulus offset—if language researchers are to understand how spoken word recognition works at all. That is, these linguistic characteristics can naturally be treated as time-series or sequential data, and we treat them as such here.

The second analysis, therefore, focuses on the time course of the predictors used in the first model by only calculating phonotactic probability, phonological neighbourhood density, and morphological complexity using the portion of the pseudoword that occurs before its uniqueness point. In this way, the statistical model captures an earlier snapshot of the recognition process than when the characteristics are calculated at the stimulus offset. If the effects trend the same between the first and second models, then this is evidence that pseudoword recognition uses the same mechanisms as real word recognition, as suggested by Luce (1986), Luce and Pisoni (1998), and Norris et al. (2000), especially since the same mechanisms are obligatorily in use before the uniqueness point. Following the fitting of the model, the results are presented and discussed.

Subsequently, a general discussion of both analyses is presented with the sort of phenomena that spoken word recognition models would need to account for to completely handle pseudoword recognition. Current models include: original cohort models (Marslen-Wilson and Tyler, 1980; Marslen-Wilson and Welsh, 1978), the Neighborhood Activation Model (Luce, 1986; Luce and Pisoni, 1998), TRACE (McClelland and Elman, 1986) and TISK You and Magnuson (2018), naive and linear discriminative learning Arnold et al. (2017); Baayen et al. (2019, 2011); Chuang et al. (2019, 2020), Shortlist B (Norris and McQueen, 2008), MERGE (Norris et al., 2000), and the cohort-like DIANA (Nenadić et al., 2018; ten Bosch et al., 2013, 2015a,b). Observations are also made about the nature of the auditory lexical decision task. The paper concludes by situating the results of the pseudoword analyses in the greater context of spoken word recognition.

Some of the main contributions of the present paper are methodological in nature. Methods by which to treat linguistic characteristics more like time-series are addressed

and discussed in Analysis 2. There is then further discussion of how the different linguistic characteristics correlate with each other (both pre-uniqueness point and at the endpoint of items). In addition, the relationship between some of these variables and the concepts they are supposed to represent is assessed in light of how the effect directions and significance levels change depending on which portion of the items is used for calculation.

2. Analysis 1

2.1. Data

For this analysis, participant responses from the aforementioned MALD data set were analysed. The data from this mega-study comprises many responses to an auditory lexical decision task. There are 232 unique monolingual native Canadian English speakers present in the MALD data set. Together, they responded to 26,800 real English words and 9,592 phonotactically legal pseudowords. There were a total of 227,179 responses in the data set, and there was a mean of 11.83 responses for each pseudoword, with a standard deviation of 1.18 responses.

The real word stimuli were selected from a variety of sources, including words from the Buckeye Corpus of Conversational Speech (Pitt et al., 2007), the CELEX database (Baayen et al., 1995), and the word list from the English Lexicon Project (Balota et al., 2007). The pseudowords were then created using the Wuggy (Keuleers and Brysbaert, 2010) program modified to work with phonological representations of words instead of orthographic representations. The parameters were set to create one pseudoword per word. Wuggy replaced one third of the sub-syllabic segments in the input word to create a pseudoword. Wuggy also maintained the length (in syllables and segments) of the input item. The stimulus items were recorded by a 28-year-old male native speaker of Western Canadian English who was trained in phonetics. The speaker was instructed to produce the words and pseudowords as naturally as possible when recording. The stimuli were then presented to participants from the University of Alberta Department of Linguistics subject pool who participated for course credit. A session of participation consisted of 400 real words and 400 pseudowords, and the order of presentation of the

stimuli in an experiment was randomised at the start of each run of the experiment. Each participant could participate in up to three sessions on separate days. More information on the data set and its creation, including a comparison of some lexical characteristics between real words and pseudowords, can be found in Tucker et al. (2019).

Using this data set allows for larger-scale analyses to be done with large sample sizes. In addition, it lets a wider variety of words and pseudowords be analysed than the monosyllabic or disyllabic items in previous studies. Overall, the stimuli present a better match for the speech that listeners are exposed to in everyday life than the single syllable stimuli often used in this type of experiment.

The data were subset so as to remove implausible responses or responses for which the variables of interest could not be calculated. Only correct responses were analysed so as to remove any responses where the listener judged the pseudoword to be a real word, potentially not using the same processing mechanisms as for other pseudowords. Initially, there were 96,049 correct responses to pseudowords in the data set. Some responses were dropped because they had transcription errors that resulted in them not being able to be parsed consistently by the guesser ($n = 7,814$, 8.86%). Initially, there were 88,235 accurate responses to pseudowords. Items were removed if they had responses faster than 500 ms ($n = 88$, 0.09%), a phonotactic probability of 0 ($n = 969$, 1.01%; these items tended to have errors in the transcriptions rather than that they were phonotactically illicit), and responses before the offset of the stimulus ($n = 1079$, 1.12%). After removing these items, 86,099 (89.64% of the initial number) responses remained to be analysed.

2.2. Linguistic predictors

The four linguistic predictors discussed above, phonotactic probability, phonological neighbourhood density, segmental uniqueness point, and morphological complexity were used in the analysis. The segmental uniqueness point is a positional measure—the phone position at which the pseudoword no longer matches any real words—as opposed to a temporal measure—the time at which the pseudoword no longer matches any real word’s phone sequence. We used the phonological neighbourhood density and

uniqueness point calculations included and described in the MALD data set in Tucker et al. (2019). Briefly, the phonological neighbourhood density for a given item was the number of items in an augmented form of the CMU Pronouncing Dictionary v0.6 (Weide, 2005) that had a Levenshtein distance of 1 from the item. As a note, we acknowledge that segmental uniqueness point does not account for when phonological processes like vowel nasalization might distinguish an item earlier than the segmental uniqueness point. However, we are choosing these variables due to their classical use in modelling psycholinguistic data, even though there are shortcomings to using them.

Phonotactic probability is not included in the MALD data set. Multiple methods for calculating phonotactic probability have been described in the literature. Vitevitch and Luce (2004) defined positional uniphone and diphone measures of phonotactic probability. Their method has been used in a number of later studies (e.g., Berent et al., 2007; Chuang et al., 2019, 2020; Gierut et al., 2010). Bailey and Hahn (2001) defined phonotactic probability as the geometric mean of the transitional probabilities between the segments of an item. The procedures for both of these methods have shortcomings, so we instead calculated phonotactic probability as a co-occurrence probability. This involves the product of the probability of occurrence of the diphones in the pseudowords. See Appendix A for a detailed discussion of the previous methods and the procedure we used for calculating phonotactic probability.

Morphological complexity for the pseudowords was operationalised as the number of possible morphological parses or decompositions the pseudoword could undergo. This particular measure was chosen over the number of morphemes because counting morphemes in pseudowords is not well defined, and the chosen measure provides more detail than a binary simplex/complex coding could. Morphemes were chosen as a unit of convenience, rather than by theoretical motivation.

To determine the number of morphological parses, a morphological guesser was developed for English. Because no orthographic representation of the pseudowords was available (and most English morphological parsers are designed to only work with English orthography), the guesser was designed to use phonological transcriptions. Note that primary stress was not indicated in the transcriptions provided in the MALD data set. As a result, stress could not be accounted for in the morphological analysis

presented here.

The guesser was built using the **Foma** finite state transducer package (Hulden, 2009). A finite state transducer is a formal way of specifying transitions between states. In the context of examining words, it was used to describe the possible sequences of prefixes, then roots, and then suffixes for an English word or pseudoword. For the present study, the finite state transducer was designed to look for all possible combinations of 0 or more prefixes, 1 root, and 0 or more suffixes, based on English prefixes and suffixes. The guesser returns all possible morphological parses that the phonological string can be broken up into, based on the prefixes and suffixes coded into it. These parses are based on potential underlying forms and not surface forms such that a pseudoword ending in [-s] would yield parses indicating a plural noun, a 3rd person singular inflection of a verb, and the possessive form of a noun. We believe this is preferable to only looking at surface level parses because there is some evidence that morphologically ambiguous real words take longer to process (Järvikivi et al., 2009; Tsang and Chen, 2013; Xiang et al., 2011). Additionally, it is impossible to predict whether a listener may gravitate towards parsing such a pseudoword as a noun or verb, or considering both types of parse, so analyzing potential underlying parses better reflects the uncertainty of the listener’s behaviour.

As an example, for the pseudoword [ab.ɪz], the guesser’s results would include (among others) [ab.ɪz] as an infinitive verb; as the non-productive prefix [ab-] (as in *obstacle* and *obstruct*) and the root [ɪz] and classed as an infinitive verb; and the noun root [ab.ɪ] with the plural suffix [-z]. Note that the root [ab.ɪ] does not seem to follow general phonotactic rules for English by ending in [ɪ]. However, the motivating idea behind these parses is that they are potential parses that a listener might consider while listening to the pseudoword. In a computational account, the listener may remove a parse from consideration on the basis that it does not follow phonotactic constraints, but processing power must still be used to remove it. And in connectionist accounts that fully-specify all possible connections between units, such forms would receive some level of activation.

All pseudowords had at least five potential parses (due to there being separate parses for interpreting the entire pseudoword as a noun or as a verb, for example), and the

number of potential parses for a pseudoword was taken as an index of its morphological complexity. Note that this measure of morphological complexity does not account for pseudo-compounding, e.g., [kãũntəɪzuvz], which could be interpreted as starting with the English word *counter*. However, the measure as proposed here should have a high correlation with one that does account for pseudo-compounds. We readily acknowledge that there are shortcomings to operationalizing morphological complexity in this fashion. It is a noisy predictor, and it includes potential parses that the listener may not actually consider, such as bases ending in vowels that are phonotactically disallowed in word-final position, such as [ɑbɪ]. However, there is no canonical method by which to morphologically parse pseudowords. Extant morphological parsers could not be used because the pseudowords were generated from phonological and not orthographic strings, and the available parsers expect orthographic strings. This variable also ignores nuances in complexity like whether an item has a prefix or a suffix, which may influence results. However, it is not trivial to devise a scheme by which to parse pseudowords in absence of any sort of sentential or phrasal context and where morpheme boundaries cannot be stated with the same level of certainty that they often can with real words. And, we do believe that this variable will generally index morphological complexity and that the trend of which items have fewer or more potential parses will align with which items have less or more complexity generally.

2.3. Results and discussion

A linear mixed-effects regression model was fit to the response time data using the `lme4` (version 1.1-28 Bates et al., 2015) package in the R statistical computing environment (version 4.1.2 R Core Team, 2021). The model fitting process followed a stepwise backwards-fitting procedure for the random effect structure and a stepwise forward-fitting procedure for the fixed effect structure. The log-likelihood test from the `anova` function in `lme4` was used to compare models. For modelling, reaction times were measured from the offset of the stimulus. In part, this decision served to factor out some of the spurious correlation between phonotactic probability from response time because longer items would have longer response times when measured from stimulus onset, and longer items have lower phonotactic probability. In a model that is

already struggling with multicollinearity, we chose to measure reaction time from stimulus offset. Choosing instead to covary the effect of length would have worsened the multicollinearity problem since length is substantially correlated with some linguistic predictors (e.g., neighbourhood density Kapatsinski, 2005).. This relationship between item length and log phonotactic probability is a result of using a product when calculating phonotactic probability. Reaction times were logged because the residuals in the model were not normally distributed when reaction time was in linear space.

In addition, we employed a version of the moving average previous response latency—as described by ten Bosch et al. (2018)—as a control predictor. Following Nenadić and Tucker (2020), we set the α parameter to 0.1 globally. The α parameter dictates how long previous trials will continue to significantly influence the moving average. At a value of 0.1, roughly 10 previous trials will significantly influence the moving average (consult ten Bosch et al., 2018).

Each predictor was centred and scaled when possible to help the model converge and put their effects on similar scales. Phonotactic probability was logged because the relationship between phonotactic probability and the response times was more linear in log space ($r = 0.26$) than linear space ($r = 0.01$). The final model had a by-subject random intercept with random slopes for trial number and moving average response latency, in addition to a by-item random intercept with a random slope for moving average response latency. Other random effect structures were attempted, but there were not enough observations in the data set to avoid a singular fit.

Control predictors such as age and sex were in the initial fixed effect structure, though no significant effects were found, and as a result, they were removed during the fitting process. We applied the model criticism procedure from Baayen and Milin (2010). We found that the residuals suggested the initial model was not predicting lower values of response time very well. As such, observations whose associated residuals in the model are at least 2.5 standard deviations away from the mean residual value were dropped, and the model was refit on the subset data. The subsequent model was noticeably less stressed and predicted lower values of response time better. A total of 1,870 or 2.17% of observations were dropped. The results of the final model after applying model criticism are presented in Table 1. Note that this model is considered

Table 1. Table of coefficients for fitted model. All predictors were centred and scaled before fitting the model.

Term	Estimate	Std. Error	t value
Intercept	6.022	0.015	409.029
Log phonotactic prob.	0.029	0.006	5.102
Neighbourhood density	0.074	0.014	5.233
Uniqueness point	0.036	0.003	13.523
Log moving average response latency	0.189	0.007	26.234
Trial number	-0.068	0.004	-16.343
Item duration	-0.193	0.004	-47.075
Number of morphological parses	0.017	0.003	6.529
Log phonotactic prob.*Neighbourhood density	-0.034	0.008	-4.051

more conservative.

The effects of each of the tested linguistic characteristics are as expected. An effect plot of uniqueness point is provided in Figure 1 to help situate the effect on the reaction times. On the whole, though, the more word-like a pseudoword is—as exemplified by it having higher values of the various linguistic variables of interest—the longer it takes the participants to reject the pseudoword. For the effects of phonological neighbourhood density and phonological uniqueness point, these effects mirror what has been observed for real words (Luce, 1986; Luce and Pisoni, 1998; Marslen-Wilson and Zwitserlood, 1989), where responses are slower for higher neighbourhood density and later uniqueness point. Similarly, pseudowords exhibiting greater morphological complexity took more time to process.

We believe that the effect of the number of potential morphological parses arises from consistent contact between the (pseudo)wordform being recognised and semantic content. Indeed, Revill et al. (2008) found that there is constant activation cascading from form to meaning during audition for real words, which must also be happening for pseudowords since listeners do not know that they are listening to a pseudoword until they have heard at least a portion of the pseudoword. The recognition process would involve any possible meaningful subcomponent of the pseudoword being paired with its corresponding semantic content throughout the listening process. Under this account, every time that [s] or [z] is heard in a position where it could be the phonetic

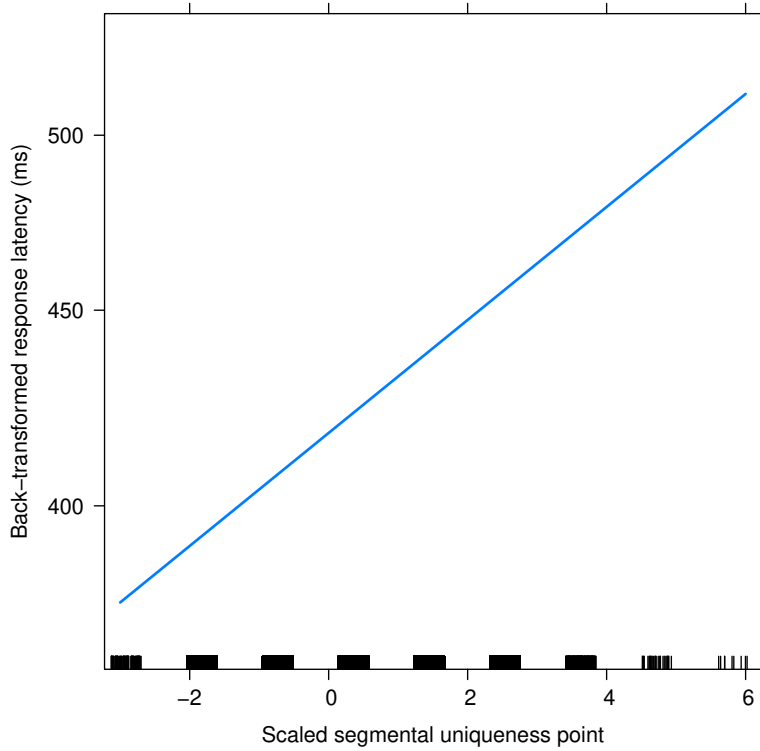


Figure 1. Effect plot for segmental uniqueness point on reaction time.

realisation of a plural marker at that point in time, the semantic content related to the English plural marker is activated. However, the activation would die off quickly because it is unlikely (to the extent of being ungrammatical) to have the plural marker *-s* occur in the middle of a word, which the listener could detect when the phonetic signal continues immediately thereafter. This is not to say that the semantic activation is substantial enough for the listener to completely experience the semantic content associated with the acoustic patterns being heard (and we could not make such a claim with our data), but rather that some degree of semantic activation occurs and impacts processing. Note, however, that the listener’s speech processing systems may also need to contend with the possibility of a word boundary within the pseudoword, e.g., that [s] or [z] might mark a word boundary, and this additional processing may also slow the listener’s responses.

The effect of phonotactic probability requires special attention to its interpretation. It was log-transformed, so the effect that was modelled was its base-*e* order of

magnitude and not its specific value. As such, the model shows evidence that when participants reject a pseudoword as not a real word, participants are sensitive to gross differences in phonotactic probability. It is possible that participants are also sensitive to more fine differences in phonotactic probability, but we do not have the appropriate evidence in our models to make such a claim.

Of all the possible two-way interactions between the tested linguistic characteristics, only the interaction between log phonotactic probability and neighbourhood density was significant. This effect is visualised in Figure 2. The effect can be bifurcated based on the number of phonological neighbours. 66.06% of pseudowords in the data set have no phonological neighbours. So, for those pseudowords, log phonotactic probability cannot vary in terms of phonological neighbourhood density. The interaction is only relevant, then, for pseudowords with phonological neighbours. It is possible that this effect is merely a statistical artefact of the presence of so many items with no phonological neighbours, which can be seen by the amount of white in Figure 2. This is not unexpected, though; data sets with many 0 values for phonological neighbourhood density are attested, with Kapatsinski (2005) reporting that 58% of the real words in the English lexicon they used had no phonological neighbours either. Note also that since phonotactic probability is strongly related to item length, this interaction could also be interpreted as one between neighbourhood density and item length.

Broadly speaking, the results match the trends seen for responses to real words, which supports our hypothesis that it is the same processes used for pseudowords as for real words. In general, the process of spoken word recognition can be thought of as convergence from the acoustic signal onto a word, as Marslen-Wilson and Tyler (1980) suggested. Should the acoustic signal diverge from all items in the lexicon, pseudoword recognition is the outcome. Specifically, we believe that it may be the process where a word form is recognised, but little to no semantic information is retrieved in relation to it, allowing listeners to make a pseudoword judgement. Word form recognition here is intended to refer to a listener having recognised that they have heard a string of speech segments that they have no semantic pairing for, that is, a sequence of speech for which insubstantial semantic activation occurred.

The first moment at which the acoustic signal can completely diverge from every

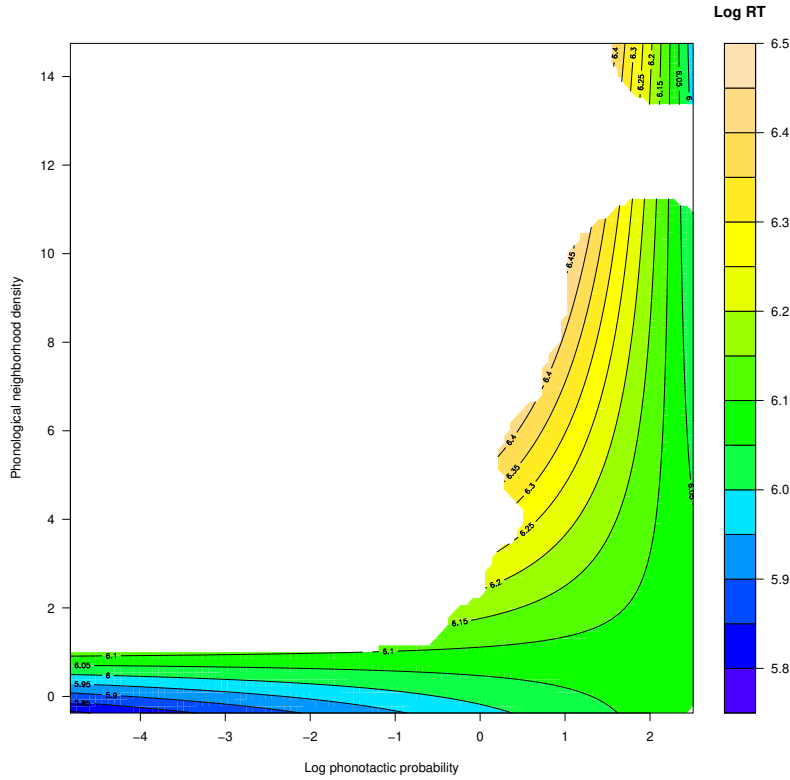


Figure 2. Contour plot of interaction between phonotactic probability and phonological neighbourhood density. Predictions for which there were no nearby observations in the data are not plotted (i.e., they are white in the plot). Blue values indicate shorter response times, whereas orange values indicate longer response times. All predictors not visualised herein were held at their mean of 0 (by centring and scaling) for numerical predictors and at their reference level for factor predictors. Note that the phonological neighbourhood density and phonotactic probability values presented in the figure are centred and scaled.

entry in the lexicon is the uniqueness point. Quite simply, a pseudoword cannot be detected until it is determined to not match any item in the lexicon. After the uniqueness point, the signal only contains negative evidence of any item in the lexicon. And yet, as noted in the data cleanup section, there were few responses recorded before stimulus offset, mirroring Ernestus and Cutler (2015). Though, Marslen-Wilson and Zwitserlood (1989) observed that the nature of auditory lexical decision requires participants to wait until the end of the stimulus to respond because a word could, at any point, have a phone shifted to make it a nonword or pseudoword, so the listener must wait until the end. Nevertheless, this observation does not apply to the pseudowords being analysed in the present study because the uniqueness point gives a definitive answer before the stimulus offset as to whether the item is a real word or a pseudoword. Ultimately, the effect of uniqueness point on pseudoword recognition raises

the question of the time course of when the studied linguistic characteristics start to truly affect processing time and whether certain kinds of information factor in sooner or later in the process.

Indeed, pseudoword recognition is not an instantaneous process. As such, it would be beneficial to know if the aspects of recognition these predictors capture have an influence across the duration of the stimulus or only after the offset. For example, can the effects of phonological neighbourhood density be seen before the end of the pseudoword? Before the uniqueness point? And, what about phonotactic probability? Are its effects notable before the uniqueness point? From the present analysis, it's not ascertainable. In the next analysis, we will address some of the temporal effects.

3. Analysis 2

The activation and competition metaphor, at its core, describes a dynamic process that unfolds over time. Various possible targets in the lexicon are vying to be recognised, and activation is continually given to targets based on how well they match the phonetic signal. We believe that the field as a whole recognises this and believes that the recognition process is dynamic. Yet, this image of activity is at odds with the way that the characteristics of a phonetic signal are calculated and analysed. Linguistic characteristics of a phonetic signal are generally calculated at the endpoint of the signal, leaving the temporal dynamics of activation, competition, and recognition to be extrapolated based on just the information contained at that one time point. In effect, the dynamics of the recognition process are estimated based only on the information a listener has at the end of hearing a word or pseudoword. For example, consider phonological neighbourhood density as it is classically calculated. It is based on whole word forms, which a listener would only have access to at the end of hearing a stimulus. But, it is plausible that competition between candidates is occurring throughout the entirety of the signal. Our goal in this analysis is to re-cast the linguistic characteristics of the phonetic signal as time-dependent.

As such, we consider each linguistic characteristic to be a time-series in this analysis, similar to how a waveform in a recording is a time-series of amplitude measurements.

To illustrate this point, consider phonological neighbourhood density as a way of counting the number of plausible competitors to a signal. The number of competitors will change over time, with activation rising and falling across large numbers of words. So, phonological neighbourhood density can be calculated incrementally as each new phone is heard to provide a coarse temporal representation of how many words are in competition at discrete time points in the signal. Each time point represents a new snapshot of the information that a listener has access to at that point in time. As this approach seemed to be not well attested in the literature, we did not examine the entire time course of each of the linguistic characteristics. Rather, we chose a simpler starting place and focused on analysing how the state of the listener’s recognition system at the moment before the uniqueness point affects the recognition and decision process. Focusing on the characteristics of certain portions of the phonetic signal has been done before. However, Vitevitch (2002) and Vitevitch et al. (2004) examined what they term “onset density”, or the proportion of phonological neighbours of a word that share the onset of the word. Yet, the calculation of the onset density still involves calculating phonological neighbours based on the offset or end state of items in the lexicon. While a listener may have access to coarticulatory information (e.g., Öhman, 1966; Whalen, 1984) over several segments, a listener would generally not have access to the end of a multisyllabic word or pseudoword they are hearing when they hear the beginning of the word, so the state of their recognition process would not necessarily reflect this information. In fact, as an anonymous reviewer pointed out, this may be part of the reason that item length is an important predictor in lexical decision since it can take longer to encounter cues that an item is a pseudoword for longer items.

Calculating a variable at various points in time and using them as predictors in this way is not unheard of in developmental studies in language acquisition (Goodman and Bates, 1997). Casting phenomena as time series data in this way is not unheard of in other fields either, such as in health sciences. For example, consider that an individual’s height, body weight, and blood pressure change over time. Earlier measurements of these variables can be used to predict future information for an individual. For example, Cook et al. (1997) used childhood blood pressure, childhood height, and

childhood weight to predict young adult blood pressure. In addition, Stovitz et al. (2008) used childhood height, childhood body mass index, and interactions thereof to predict adult body mass index. The same logic of using variables calculated at previous time points to predict future behaviour undergirded the present analysis of using the linguistic characteristics of the signal just before the uniqueness point. Our decision to use the uniqueness point as the cutoff was predicated on the thought that the uniqueness point is the first moment when the pseudoword can be identified and decided upon. This makes the uniqueness point a meaningful time point at which to recalculate the linguistic characteristics of the signal. The re-calculated linguistic characteristics thus represent a snapshot of the information the listener would have just before the uniqueness point is reached.

Survival analysis, or other related methods, could also be used to analyse the time-course effect of the predictors as well. However, survival analysis uses static versions of the predictors in modelling the dynamics of their effect on the response variable. We are interested in how the dynamic nature of the linguistic predictors could affect the participants' response latency. This dynamic nature is of particular interest because participants would not have access to much of the information that predictors like neighbourhood density represent until stimulus offset. It is thus unclear for the moment how the linguistic predictors would need to be calculated for survival analysis to effectively analyse the time-course of their effects on response latencies.

Since phonological neighbourhood density, log phonotactic probability, and the number of morphological parses were significant predictors when looking at the entirety of the stimuli, they should be predictive before the uniqueness point as well. However, it would certainly be possible to calculate these predictors at other points during the course of hearing the pseudoword as well, such as after the first phone. Note that, of all the pseudowords present in the MALD item data set ($n = 9592$), 16.59% ($n = 1591$) have their uniqueness point after the last phone, meaning that there are real words that begin with the same phone string but that the phone string itself is not a word. Some examples include the pseudoword [æf], which includes the same onset as *after*, and the pseudoword [twi], which includes the same onset as *tweak*. This is, of course, natural for real words too, since *cat* [kæt] contains the same onset as *cats* [kæts] and

catastrophic [kætəstɹɪfɪk].

To calculate the phonological neighbourhood density for these segments of pseudowords, the Levenshtein distance was calculated between the non-unique portions of each pseudoword (up to but not including the phone indicated by the positional uniqueness point in the data set) and each entry in the augmented CMU Pronouncing Dictionary. The dictionary entries were truncated to be the same length as the non-unique portion of the pseudoword if dictionary entries were longer than the non-unique portion of the pseudoword. The number of comparisons that had a Levenshtein distance of 1 was taken to be the pseudoword’s phonological neighbourhood density before the uniqueness point. The phonological neighbourhood density values were overall higher when calculated on these sub-portions of the items ($M = 108.51$, $SD = 384.89$) in comparison to values from the previous analysis ($M = 5.12$, $SD = 13.69$).

Phonotactic probability was calculated by only considering the probabilities of the diphones of the non-unique portion of each pseudoword, with no offset token. These values were also higher when calculated in this fashion ($M = 4.0 \times 10^{-9}$, $SD = 9.3 \times 10^{-8}$) when compared to the original formulation from Analysis 1 ($M = 7.5 \times 10^{-11}$, $SD = 7.4 \times 10^{-9}$).

The number of potential morphological parses for the non-unique portion of a pseudoword was calculated similarly to phonotactic probability. The non-unique portion of each pseudoword was run through the finite state transducer from before. Then, the number of guesses made for the structure of the pseudoword was taken as the number of potential parses. The average number of parses for the non-unique portion was slightly smaller ($M = 6.82$, $SD = 4.05$) compared to the number of parses for the entirety of each pseudoword ($M = 9.348$, $SD = 8.29$). This is to be expected because there are fewer possible phone combinations for shorter strings, which would deflate the number of parses that could be performed. That is, the average number of parses is related to the number of possible substrings of a string, and longer strings always have a greater number of possible substrings than a shorter string.

For a control variable similar to item duration, the temporal uniqueness point was calculated for each pseudoword as an indication of how much time had elapsed by the time the positional uniqueness point occurred. Finally, positional uniqueness point

was not used in the modelling process. Even though uniqueness point has a strong relationship to response latency, it is incompatible with this type of analysis because it is information that the listener, by definition, will not have access to.

As before, all applicable variables were centred and scaled. Observations where the response time was before the offset or 500 ms were dropped, leaving 87,033 responses (90.61% of the original correct responses) to analyse. No responses needed to be dropped for the reason of having a phonotactic probability of 0 (which cannot be meaningfully logged). Phonological uniqueness point was not present in the model because it is information the listener would not have at the time point being analysed. Distribution plots comparing these pre-uniqueness point predictors to their full-length counterparts can be found in the supplementary materials. Log neighbourhood density plots are also provided because the linear space version of the phonological neighbourhood density plots are difficult to interpret due to outliers.

3.1. Results

The fitting process was the same as for the previous analysis. The resulting model had log phonotactic probability, neighbourhood density, number of morphological parses, temporal uniqueness point, log moving average response latency, and trial number as fixed-effects. All two-way interactions between the phonological uniqueness point, log phonotactic probability, and the number of potential morphological parses were checked. Those interactions that did not contribute to the model fit during the fixed-effect fitting procedure were dropped, leaving only the interaction for log phonotactic probability by number of morphological parses. The model also had a by-subject random intercept with a random slope for moving average response latency and a by-item random intercept with a random slope for moving average response latency. We attempted to fit random slopes for trial number, but the `rePCA` function from the `lme4` package suggested that the random slope was not adding much explanatory power, so we opted for the more conservative model that did not include the random slopes for trial number. This model was subjected to model criticism as in analysis 1. The original model was also not predicting lower values of reaction time well, as indicated by the residuals; the model resulting from the model criticism procedure predicted the lower

values far better. The results of the more conservative model from the model criticism procedure on the final model are presented in Table 2. Note that the reported interaction is insignificant. It was kept in the model because it was significant before adding the random slopes for the moving average response latency, but it was insignificant afterwards. Notably, the effect of neighbourhood density was not significant either.

Table 2. Table of coefficients for the model fitted with the pre-uniqueness point data. All predictors were centred and scaled before fitting the model.

Term	Estimate	Std. Error	t value
Intercept	6.027	0.013	451.360
Log phonotactic prob.	0.018	0.004	4.436
Trial number	-0.055	0.002	-29.875
Neighbourhood density	0.012	0.009	1.254
Log moving average response latency	0.195	0.006	31.087
Temporal uniqueness point	0.054	0.004	14.955
N. morph. parses	-0.016	0.003	-4.773
Log phonotactic prob.*N. morph parses	-0.008	0.004	-1.948

We subsequently performed a correlation analysis to verify that the regression using the pre-uniqueness point linguistic variables was representing information that was different from the previous model with linguistic variables calculated for the full item. We further subset the data set used for the current analysis so that no items with a full-length phonotactic probability of 0 were in the data set, thereby dropping an additional 934 data points (1.07% of the data used in the regression models). A corrgram visualising the correlations between the variables, which Tomaschek et al. (2018) suggested as a diagnostic for multicollinearity, is presented in Figure 3.

The correlations between the full-length variables and their pre-uniqueness point counterparts were negligible, with the exception of the number of morphological parses. This is to be expected, however, based on the method in which the number of morphological parses was calculated. Specifically, the number of possible parses by the end of the word must be greater than or equal to the number of possible parses by the uniqueness point. And, because there will very likely be more parses when the parser analyses potential suffixes, it is reasonable and unsurprising that a high number of morphological parses before the uniqueness point is predictive of a high(er) number

Lg pre ph pr	Pre up ND	Temp UP	Pre m parse	Lg ph pr	ND	UP	M parse	N phones
0.32	-0.45	-0.21	0.18	-0.14				
-0.22	-0.12	0.35	-0.10	0.53				
0.14	0.03	0.04	0.27	-0.19	-0.18			
0.15	0.06	0.47	0.71	-0.17	-0.12	0.13		
-0.84	-0.38	0.10	0.18	-0.96	-0.52	0.31	0.20	
-0.09	-0.04	-0.31						
-0.20	-0.08							

Figure 3. Corrgram displaying the correlations between the linguistic variables used in analyses 1 and 2. The color of the number in each cell is an indication of the magnitude of the correlation. The more red, the greater the magnitude of the correlation, and the more blue, the smaller the magnitude of the correlation. Each column heading also serves as a row heading for all values to the left. "Lg pre ph pr" is the log pre-uniqueness point phonotactic probability, "Pre up ND" is the pre-uniqueness point neighbourhood density, "Temp UP" is the temporal uniqueness point, "Pre m parse" is the number of pre-uniqueness point morphological parses, "Lg ph pr" is the log phonotactic probability of the full item, "ND" is the neighbourhood density of the full item, "UP" is the uniqueness point of the item, "M parse" is the number of morphological parses for the full item, and "N phones" is the number of phones in the full item.

of parses for the full item, and vice-versa. To verify that this switch in direction was not due to removing uniqueness point from the model as a control for item length, we fitted an additional model that also had the positional uniqueness point variable in it, applying model criticism to it as before. Neither the trend direction nor the level of significance for the number of morphological parses changed. The table of coefficients for this additional model can be seen in Table S1 in the supplemental materials. Many of the other effects changed direction in the model—likely due to problems with multicollinearity. Since this additional model was fitted strictly to check the effect trend of the number of morphological parses, it was not interpreted further.

There is a high negative correlation between log pre-uniqueness point phonotactic probability and full-item uniqueness point. We are confident that this is due to the manner in which we calculated phonotactic probability, where each successive diphone would decrease the probability by a similar order of magnitude. As such, cutting the probability calculation off at the phone before the uniqueness point should make the actual uniqueness point readily predictable from the log phonotactic probability. In the grander scheme, we believe this result speaks to a weakness inherent in phonotactic probability as a predictor since a lot of its effect may be reducible to some form of the number of phones, diphones, or some other unit in the word.

We performed an additional analysis on the pre-uniqueness point and full-item

neighbourhood density values. This analysis was to allay any doubt that the low correlation between the pre-uniqueness point neighbourhood density and full-item neighbourhood density values was due merely to the number of full-item neighbourhood densities that had a value of 0. Consequently, we calculated the correlation between only those responses to items that had a phonological neighbourhood density value greater than 0. No pre-uniqueness point neighbourhood density values were less than 1, so the subsetting process only involved the full-item neighbourhood density values. From the 87,033 responses in the data used in the modelling, 57,644 responses were dropped (66.23% of the responses used in the regression analysis), leaving 29,389 responses. The resulting correlation coefficient was $r = 0.12$.

Most of the linguistic characteristics were significant in the model, with the exception of neighbourhood density. Based on the correlation analysis, it seems that the effect of log phonotactic probability may simply be due to its correlation with uniqueness point. Both the insignificance of neighbourhood density and the correlation of log phonotactic probability with uniqueness point suggest that these characteristics are not good proxies for the dynamics of competition before the uniqueness point.

Regarding neighbourhood density, there are two possibilities. The first possibility is that it is a poor proxy for competition. While there is a wealth of literature reporting its effects when calculated for full-length words and pseudowords, it may not generalise well to being calculated at other points of the pseudoword. The second possibility is that the sort of competition effects neighbourhood density represents arise after the offset of the item. Either of these possibilities suggests that the nature of competition is more complex than neighbourhood density would suggest. We do not believe this is a novel conclusion, and, in fact, it seems that some researchers likely already believe this (Kapatsinski, 2005; Nelson and Wedel, 2017; Vitevitch and Luce, 2016). Yet, continued usage of just the one-edit definition (Artiunian and Lopukhina, 2020; Chuang et al., 2020; Diaz et al., 2021) of phonological neighbours fails to accord with prevailing thoughts on lexical competition.

The effect of log phonotactic probability suggests that the likelihood of the set of diphones that occur before the uniqueness point in a pseudoword already has a tangible effect on response latencies. That is, more likely sequences are making the recognition

process take more time even by the time the uniqueness point is reached. The effect is overall similar to the effect phonotactic probability had when calculated at the end of the word in Analysis 1. That being said, because of its high correlation with uniqueness point, the results of the regression model are unclear. The model cannot contain both log phonotactic probability and uniqueness point as predictors and still be interpretable, so the effect of log phonotactic probability cannot be reasonably interpreted as separate from the effect of uniqueness point. And, because a monotonic sum is proportional to the number of items being summed—where a sum is equal to the number of items being summed times the arithmetic mean of the items—it is actually unclear whether log phonotactic probability and string length are conceptually different. It is beyond the scope of this study to determine which concept has primacy over the other, if either does.

The effect of the number of potential morphological parses of a pseudoword has changed direction in comparison to the first analysis. It suggests that there is a categorical shift in behaviour at the uniqueness point. Before the uniqueness point, the number of potential morphological parses makes it take less time to recognise a pseudoword. While, after the uniqueness point, it takes more time to recognise a pseudoword. For the moment, we do not have enough evidence to disconfirm any hypotheses about this behaviour. It could be that the categorical shift has to do with a change in behaviour once the recognition system realises it is no longer hearing a real word. It could be that this manner of quantifying morphological complexity does not generalise very well across experiments and analyses. Or it could be that real words also exhibit this behaviour. Evidence is needed on how real words behave to determine whether the categorical shift could plausibly be due to the recognition system shifting from recognising a real word to recognising a pseudoword.

Finally, the temporal uniqueness point is an indication of the duration of the pseudoword before the uniqueness point. Its effect merely describes how the more time it takes to get to a pseudoword’s uniqueness point, the longer it takes to respond to. And, this result is intuitive; the item can’t be identified as a pseudoword until after the uniqueness point, so having to wait longer for that point means a response will take longer. Note that this effect is the reverse of the effect of duration in the previous

model, where longer items were faster to respond to. This discrepancy is expected because we measured the response time from stimulus offset in both analyses. Longer item durations allow for more processing time before stimulus offset, meaning that response latencies should be slower. Whereas, for the aforementioned reason, a later temporal uniqueness point should indicate longer response times.

Overall, the results of this analysis are inconclusive, at best. Calculated before the uniqueness point, the effect of phonological neighbourhood density was insignificant, and the effect of log phonotactic probability could be reduced to the effect of phonological uniqueness point. The effect of the number of potential morphological parses changed in direction, though, which could indicate a shift in the cognitive processing techniques at play while the participant is listening to the pseudoword.

Because spoken word recognition is widely thought to be dynamic, there is a strong reason to believe that there should be a way to treat the number of competitors as a time series and make predictions about how the number of competitors at a certain time point affects response latency in lexical decision. However, phonological neighbourhood density was not generalisable to this sort of analysis. For neighbourhood density, we believe this is because the one-edit definition of a neighbour is not a good indicator for being neighbours, which Kapatsinski (2005) also argued. Even Luce (1986) highlights that this method of determining whether items are neighbours is more of a convenience method and should eventually be replaced. We suspect quantifying competition on the basis of acoustic distance like in Kelley and Tucker (2022) would yield results which better correspond to actual distance. Per a comment from an anonymous reviewer, it may also prove fruitful to weight the differences between segments by using distinctive features, given some evidence that they relate to cohort sizes (Kotzor et al., 2017; Lahiri, 2018; Lahiri and Reetz, 2010) and that some featural changes matter more than others (Connine et al., 1991).

4. General discussion

The results from the present study indicate that linguistic characteristics which have previously been shown to affect the processing of real words also affect the recognition

of pseudowords. However, some of these linguistic characteristics were not robustly enough related to actual cognitive processing that they could be generalised to model the state of lexical competition at an earlier point in the word, even though there must be a dynamic, time-bound nature to how pseudowords are recognised. These findings have implications for models of spoken word recognition, as well as how auditory lexical decision data should be interpreted.

The notion of recognising a pseudoword has historically been nebulously defined. The closest descriptions of this process comes from the aforementioned neighborhood activation model (Luce and Pisoni, 1998) and the Merge model (Norris et al., 2000). In the neighbourhood activation model, a failure to activate a lexical candidate sufficiently results in the identification of a pseudoword. Similarly, in Merge, reaching a deadline for no lexical entry receiving enough activation will prompt pseudoword recognition. To reiterate the conceptualisation of spoken word recognition from our analyses, there are ultimately two extreme outcomes of spoken word recognition: the signal converges on a lexical item, or the signal diverges from all lexical items. This description is also similar to the convergence description Marslen-Wilson and Tyler (1980) gave. Pseudoword recognition is the end result of divergence from all items in the lexicon. In effect, then, the default behaviour of the recognition system is to recognise real words, and the end result of the signal diverging may require suppressing this behaviour. And, when the incoming signal is more word-like with, for example, high phonotactic probability, the suppression would require effort (and thus takes more time).

It is unclear from our data or previous data, though, what exactly pseudoword “recognition” is. It may be that the general cognitive system has a distinct state it can be in when some linguistic input has not matched any entries in the lexicon. Or, it is possible that the “recognition” state is simply a form of nothingness or inactivity due to no lexical entry being sufficiently activated. Whatever it is, though, following recognition, a listener in an auditory lexical decision task must then proceed to their decision process, a point which we will take up again shortly.

The meanings of the lexical predictors for pseudowords are united under the convergence/divergence idea. Phonological neighbourhood density is, instead of the number

of competitors, the number of items that the signal shows some level of convergence toward. The uniqueness point marks the moment when the signal begins to diverge overall from the items in the lexicon. The number of morphological parses is an operationalisation of what potential smaller units of meaning there are in the pseudoword, which is related to how much semantic activation could occur while listening to the signal. And, phonotactic probability is a reflection of how well the acoustic patterns in the data match a listener's expectations based on the phonotactics of their language and how many items may remain in competition at a given point in time; when the acoustic patterns don't match the phonotactics, there should be greater divergence and fewer potential matches, which facilitates pseudoword recognition.

The idea that items are processed at various levels is not new. As previously mentioned, Vitevitch and Luce (1998), for example, posit that words are processed at both a lower, sublexical level, where phonotactic probability plays a role, and at a lexical level where phonological neighbourhood density plays a role. However, it is possible that these different levels of processing are epiphenomena of the processing system, where longer-term activation patterns—like those for whole words—that yield semantic activation can be observed as “higher-level” effects, and shorter-term activation patterns—like those for affixes—that yield little to no semantic activation can be observed as “lower-level” effects. While it is possible to argue for the existence of classical hierarchical processing from this concept—e.g., phonological, morphological, and syntactic levels of processing—doing so would situate them as nothing more than metaphorical summaries of acoustic activation patterns.

Naturally, these ideas lead to a need to reconsider the broader picture of spoken word recognition. Pseudoword recognition has some demonstrably systematic aspects to it, based on the results from our analyses. And this point makes sense in light of the fact that humans can learn new words. After all, what is a pseudoword but a new, decontextualised word that someone hasn't told you the meaning of? Then, pseudoword recognition seems to be a necessary—though not sufficient—task for word learning.

In sum, it is apparent that the recognition of pseudowords uses the same processes as real words. As such, models of spoken word recognition or participant behaviour

in auditory lexical decision tasks that do not have some sort of pseudoword recognition mechanism are necessarily incomplete. This incompleteness parallels neglecting to implement conditional behaviour in the reverse-engineering of an algorithm based on inputs and outputs. To be sure, though, the “output” for spoken word recognition is not well-defined enough to say for sure what kind of modelling or algorithm should be proposed. Yet, it is obvious based on our results and previous results that cognitive processing goes on when the mind handles pseudowords, and they must be handled in a similar way as real words because cognitive proxies like phonological neighbourhood density and uniqueness point are relevant for real words and pseudowords alike.

Following the algorithm reverse-engineering analogue, we believe there are a number of aspects of behaviour regarding pseudowords that can be described. Any model of spoken word recognition that seeks to be complete must account for these aspects.

First and principal among these aspects is that a model of spoken word recognition must explicitly lay out how pseudowords are handled by the recognition system. Or else, the system must be general enough that pseudowords are already handled by the available mechanisms. There is a more open question of whether it is more advisable to have pseudoword recognition incorporated as a possible outcome of the system or if pseudoword recognition is better handled with a *post hoc* decision mechanism, or some combination thereof. Nevertheless, it is clearly necessary for complete models of spoken word recognition to describe what happens when there is no suitable recognition candidate for the incoming signal.

Second, a complete model of spoken word recognition must allow the number of real word competitors to influence the cognitive processing of pseudoword stimuli. The results of our first analysis make this point very clearly because phonological neighbourhood density is significant. While the results from our second analysis add to evidence that it is an inadequate proxy for the number of competitors, phonological neighbourhood density is at least a noisy index of how much competition is occurring during the audition process for a particular item. Perhaps the most well-known example of a model of spoken word recognition that allows the number of competitors to influence the recognition results is the Neighborhood Activation Model (Luce and Pisoni, 1998). However, the DIANA model (ten Bosch et al., 2015a,b) is far more

mathematically and computationally explicit about how the number of competitors influences the recognition and decision process.

Third, a complete model of spoken word recognition must process meaningful chunks of an item as it is being processed. The necessity of this behaviour is shown by the effect of the number of potential morphological parses in the first and second analyses. We do not believe that this specifically means morphemes and affixes, though our operationalisation of morphological complexity assuredly was built on those notions out of convenience. The naive discriminative learning (Baayen et al., 2011) and linear discriminative learning (Baayen et al., 2019) frameworks have modelled morphological effects, even without explicitly coding them into the models. Applications of the techniques to spoken word recognition (e.g., Arnold et al., 2017; Chuang et al., 2020; Shafaei-Bajestan et al., 2020) should also exemplify this kind of behaviour. However, because naive discriminative learning and linear discriminative learning are simply the application of solving a least-squares problem in linear algebra, the effects of morphological complexity could even be explicitly coded into specific models as one of the variables in a matrix.

It is difficult to say *a priori* whether any model already does account for these phenomena. It is possible that an inspection of the code or math involved in such a model may reveal its sensitivity to some or all of these phenomena, it would ultimately be more fruitful to actually use the models in computational simulations, as in Nenadić and Tucker (2020). There are, however, two difficulties in performing these simulations. The first is that many models lack easily accessible computational implementations, such as the neighbourhood activation model—for which the PARSYN implementation does exist (Luce et al., 2000) but was not released with its associated publication. Second, few models use acoustic information, so they only relate so much to *spoken* word recognition. Notable exceptions are DIANA, some naive/linear discrimination models, EARSHOT (Magnuson et al., 2020), and Fine-Tracker (Scharenborg, 2010).

Ultimately, we do not believe that any of the models currently available seem equipped at this moment to handle pseudowords. While pseudowords in and of themselves are not necessarily the most ecological kind of data to account for, they are analogous to a novel real word that has been encountered for the first time. What all

the models discussed seem to lack, then, is an in-built mechanism to account for the acquisition of new vocabulary. Certainly, it is feasible to extend some of them, and merely recognising pseudowords as we have outlined here may indeed be a good first step toward building the capacity in the models to handle new items in the lexicon. Currently, this remains an area of growth for these models, however.

Finally, a note on the auditory lexical decision process as it relates to the idea of the signal converging or diverging on lexical items. Based on the effect of morphological complexity in both regression models, the pseudowords could be said to be making contact with semantic information when possible, as Chuang et al. (2019) and Chuang et al. (2020) also found. Or rather, it could be said that meaningful units are activated whenever the acoustic information matches them, whenever possible. There is a general affinity between this idea and results from Vroomen and de Gelder (1997) where words embedded within other words effected priming in certain contexts. In this understanding, the process of spoken word recognition becomes a continuous process, as opposed to a discrete one. And, this is exactly what the activation and competition metaphor suggests. When hearing a real word, substantial semantic patterns are activated over time, corresponding to the times that a listener has heard that real word. A real word represents then a longer-term acoustic pattern than an affix, for example. And, indeed, the convergence process has the potential to be modelled using some sort of acoustic distance over time from the input signal to all the items in the lexicon. The accumulation of acoustic distance would suggest divergence from an item over time, while having little acoustic distance accumulating over time would suggest the signal converging on a lexical item. But, that remains for future work.

Per some insightful comments from L. Boves (personal communication, Sep. 23, 2020), there are two other distinct decision strategies that participants may use during lexical decision. The first of these is that a participant may judge that the stimulus is too acoustically distant from any word in their lexicon and reject the item for that reason. And, acoustic distance has indeed been found to relate to word recognition in auditory lexical decision, where greater average distance from a word to the lexicon facilitates responses and vice-versa (Kelley and Tucker, 2022). The second of these is that there may be some features of the stimulus that clue the listener in that they

are hearing a pseudoword. Ford et al. (2018), for example, found that shorter phone durations facilitated word recognition in auditory lexical decision. It seems reasonable that the inverse may be true as well, that longer phone durations facilitate pseudoword recognition, which would give the listener some sort of stimulus feature to latch onto when rendering their decision in the lexical decision task. The decision process is likely a complex conjoint of these processes and others, and it should not be neglected when creating models of spoken word recognition.

What is currently unknown is how little semantic information must be contacted, how much acoustic distance must accumulate, or how prevalent pseudoword-specific features must be before a pseudoword is judged to have been recognised, let alone how these aspects should be measured and quantified. That being said, there is reason to suspect that virtually any disruption at the segmental or higher level can render the semantic activation insubstantial. Otherwise, changing a single segment in a word would not have the ability to render it a pseudoword. Future research in spoken word recognition should work to describe the nature of the convergence/divergence process of spoken word recognition and where a gradient or discrete threshold may exist for participants determining whether they have heard a pseudoword or not.

In light of the results of the pre-uniqueness point analysis, future research on this topic should investigate the variables that are commonly used as predictors in psycholinguistic experiments and whether they are truly distinguishable from each other. Consider log phonotactic probability, which correlated at a very high level with item length. If phonotactics is supposed to come to bear on the processing of pseudowords, perhaps an alternative operationalisation of phonotactics should be found that does not correlate so highly with item length, such as factorisation into mean diphone probability across the item and item length. We similarly acknowledge that our operationalisation of morphological complexity is simplistic and coarse and likely misses much of the variability in responses expected with a more granular measure of morphological structures. We additionally hope that the MALD pseudoword data is used as the basis for more sophisticated analyses of the effects of morphology (and other linguistic phenomena, including wordlikeness) on pseudoword recognition in the future.

5. Conclusion

The present study aimed to provide a detailed description of pseudoword processing. The focus was on ascertaining whether or not linguistic characteristics that have been found to be predictive for word processing are also predictive for pseudoword processing. In each case, the linguistic characteristics had significant effects in regression models of response time in auditory lexical decision tasks. These effects were significant whether the predictors were calculated at the uniqueness point of a pseudoword or at the end of a pseudoword. Such findings indicate that lexical characteristics of a pseudoword could provide predictive ability even before the pseudoword could be identified as not belonging to the lexicon, suggesting that the discrimination process is ongoing throughout the time course of the signal a listener is experiencing. Moreover, lexical processing seems to continue after the uniqueness point occurs, suggesting that the uniqueness point is not as important to recognition as cohort-like models of spoken word recognition make it out to be. From these results, a description of the spoken word recognition process was offered whereby the recognition occurs when the audio signal converges on items in the lexicon, while a failure to recognise a word occurs when the signal diverges from all items in the lexicon.

Many spoken word recognition models proposed over the years generally do not handle pseudowords particularly well. Framing pseudoword recognition as the divergence of the acoustic signal from items in the lexicon requires that such models account for the recognition of pseudowords, as it is a natural outcome of the recognition process. Going forward, pseudowords provide a promising landscape for investigating what happens when the audio signal does not converge on any items in the listener's lexicon. Additionally, pseudowords can be confidently used to investigate the effects of lexical characteristics in a controlled manner. Finally, pseudowords should be thought of as more than mere distractors in experiments, since they involve the same processing mechanisms as real words.

Acknowledgements

The authors would like to thank the attendees of the NorthWest Phonetics & Phonology Conference 2017, the Acoustics '17 Boston Conference, and the Western Conference on Linguistics 2017 for their feedback on various versions of this project. We would also like to thank Lou Boves and three anonymous reviewers for their comments on this manuscript. This project was funded by a Social Sciences and Humanities Research Council Grant #435-2014-0678.

Disclosure statement

No potential conflict of interest was reported by the authors.

Appendix A. Phonotactic probability

Vitevitch and Luce (2004) define phonotactic probability positionally, giving options for both uniphone- and diphone-based measures. The lexical frequency information comes from Kučera and Francis (1967), while drawing the phonetic transcriptions and words from an online version of the Merriam-Webster Pocket Dictionary. For the uniphone positional measure, and for each particular phone and position pairing possible in the lexicon, the sum of the logged frequency counts of all the words containing that pairing was divided by the sum of the logged frequency counts of all words with that position available for a phone. Formally, this is given in Equation A1:

$$p(s, i) = \frac{\sum_{\sigma \in S_i} \log_{10} f_{\sigma}}{\sum_{\omega \in W_i} \log_{10} f_{\omega}}, \quad (\text{A1})$$

where $p(s, i)$ is the probability of segment s at position i , S_i is the set of all words containing s at position i and W_i is the set of all words containing any segment at position i (that is, all words that have at least i segments). For a given item, then, its phonotactic probability is taken as the sum of this calculation for each of its segments,

accounting for their positions. The function for calculating the phonotactic probability of a word, $pp(w)$ is given formally in Equation A2:

$$pp(w) = \sum_{(s,i) \in w} p(s,i), \quad (\text{A2})$$

where (s,i) is a pair containing a segment s of word w and the position i where s occurs in w . The diphone version was calculated analogously, only they used diphones instead of single phones.

Vitevitch and Luce (2004) claim that the use of the log function helps the measure better represent human sensitivity to log frequency effects. However, some characteristics of this definition of phonotactic probability are undesirable. Principally, taking the log transform of count data before performing the division operation, as in Equation A1, makes it more difficult to interpret the output in a well-defined manner. This is perhaps easier to see when expressed in an equivalent manner, as in Equation A3, which is clearly not recognisable as a traditional probability value. It may well be that the result is a good predictor of participant behaviour, but it can't reasonably be conceived of as a probability value to represent phonotactic probability.

$$p(s,i) = \log_{10} \left(\left(\prod_{\sigma \in S_i} f_{\sigma} \right)^{\frac{1}{\prod_{\omega \in W_i} f_{\omega}}} \right) \quad (\text{A3})$$

There are other methods that can be used to account for the concern that logged values better reflect human perception. One example is to calculate the probability based on count data first and then log that probability value. This is what is known as “log probability”, and it can be easily mapped back to a standard probability value between 0 and 1. It is also more transparent in terms of what it represents about the count data.

An additional concern is that words that have a frequency count of 1 will not come

to affect the probability values for any sequence because $\log(1) = 0$. Vitevitch and Luce (2004) also do not state how they account for items with a frequency of 0, for which the log function is undefined. Finally, while 0 is the lower bound for their method of calculating phonotactic probability for a word, there is no theoretical upper bound. As such, a word or pseudoword could be assigned a phonotactic probability value greater than 1, violating the definition of classical probability. Consider counting the beginning of a word as a phone as a simple example demonstrating this property. For example, the word *cat* would be represented as $/\#kæt/$ with the “#” representing the beginning or onset of a word. All words would have such a beginning symbol, so the numerator and denominator in the fraction defining phonotactic probability are equal when considering $/\#/$ in the first position, thus its calculated value of occurring at the first position is 1. Having even one phone following the $/\#/$ with a non-zero value will yield an overall value greater than 1. Whether in practice such values are often observed remains to be seen. But, it is nevertheless difficult to argue that these values can be interpreted as a proper probability if it is even possible for values greater than 1 to be obtained.

Bailey and Hahn (2001) used a similar approach to Vitevitch and Luce (2004) without the log functions to calculate transitional probabilities for diphones and triphones, though the source of the frequency counts for this metric does not seem to be mentioned in their paper. For the composite word scores, they took the geometric mean of the conditional probability scores of the segments that make up the word to calculate the score for the word. By calculating the geometric mean instead of the arithmetic mean, a true probability value is initially calculated, but it becomes less clear what the value is once it is raised to the power of $\frac{1}{n}$ to finish calculating the geometric mean. Janse and Newman (2013) used a similar method involving a mean, though CELEX (Baayen et al., 1995) was used for frequency counts. This manner of calculating phonotactic probability will converge toward a value for the word or item in question, though it is unclear what this value would be or represent. Adding subsequent segments would not necessarily drive the probability of the sequence down, which does not match the intuition that a word or pseudoword consisting of, say, 500 segments is improbable.

For the purposes of this study, we operationalised phonotactic probability as the

probability that a particular sequence of diphones would co-occur, based on the relative frequency counts of each diphone in the language. We made this decision based on the results from Bailey and Hahn (2001), in that diphones seemed the least complex unit to achieve the greatest predictive power. They also stated that diphone treatments of phonotactic probability are the most common. As well, Pierrehumbert (2003) claimed that triphones are difficult to learn in comparison to diphones, so diphones seem the best choice for predictive power and closeness to speaker knowledge. The idea of using a co-occurrence probability, which is calculated with a product like Coleman and Pierrehumbert (1997) do, is not new. Yet, because previous and popular methods of calculating and defining phonotactic probability have not done this, we believe it worth being explicit about this choice.

We used the same augmented CMU Pronouncing Dictionary version 0.6 (Weide, 2005) used by Tucker et al. (2019), as well as COCA in our calculations. We began by finding the overall frequency-based probability of occurrence for each diphone found in the CMU Pronouncing dictionary. The frequency of each diphone was calculated by taking each occurrence of it in the CMU Pronouncing Dictionary, multiplying it by the frequency of the word in COCA, and adding the resulting product to the tally of occurrences of the diphone in question. This results in a token frequency instead of a type frequency. Richtsmeier suggested using type frequencies instead, but they correlated at a value of approximately 0.99, so we don't believe there would be much of a difference. Word onset and word offset were considered phones, such that *cat*, for example, would be processed as onset+k, kæ, æt, t+offset. Words that did not occur in both datasets were dropped. The probability of occurrence of each diphone was calculated as the diphone's frequency divided by the total count of all diphones observed. This process is given formally in Equation A4

$$p(s) = \frac{\sum_{\sigma \in S} f_{\sigma}}{F}, \quad (\text{A4})$$

where S is the set of all words containing the diphone s , f_{σ} is the frequency of a word σ containing diphone s , and F is the number of diphones in a word times the word's

frequency, summed over all words occurring in both the CMU Pronouncing Dictionary and COCA. Effectively, F is the total number of diphones observed in the subset of COCA words that have pronunciations in the CMU Pronouncing Dictionary.

We then took the phonotactic probability of a pseudoword to be the product of the probabilities of occurrence of each diphone in the pseudoword, which is what Vitevitch and Luce (2004) and Bailey and Hahn (2001) refer to as the co-occurrence probability. Formally, our function for calculating the phonotactic probability of a word $pp(w)$ is given in Equation A5:

$$pp(w) = \prod_{s \in w} p(s). \quad (\text{A5})$$

Note that defining phonotactic probability as a product instead of a mean or summation of pseudo-probabilities has a few important properties. The first among them is that it concentrates the information revealed about the phonotactic probability of a sequence at the beginning. While multiplication is often commutative and associative, there is a natural given order in which to carry out the operations here, that being the order in which the diphones occur in the pseudoword. And, the rate at which the probability converges toward 0 will slow down as later and later terms are encountered. Analogously, the first few segments in a word or pseudoword are likely where the most discriminative information would be contained. This is due to the fact that the number of good possible matches for what’s being heard decreases quickly at the start of the sequence and slowly at the end of the sequence, with the largest decreases happening upon hearing the first few phones. Second, the probability converges asymptotically toward 0 for sufficiently long sequences of segments, matching the linguistic intuition that a sequence of, for example, 500 phones is an improbable occurrence for a word in a language. The implementation of this method of calculating phonotactic probability used in this study is available in the `Phonetics.jl` package (Kelley, 2020) for the **Julia** programming language (Bezanson et al., 2017).

References

- Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, 38(4):419–439.
- Arnold, D., Tomaschek, F., Sering, K., Lopez, F., and Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PloS one*, 12(4):e0174623.
- Artiunian, V. and Lopukhina, A. (2020). The Effects of Phonological Neighborhood Density in Childhood Word Production and Recognition in Russian Are Opposite to English. *Journal of Child Language*, 47(6):1244–1262. Publisher: Cambridge University Press.
- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019.
- Baayen, R. H. and Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2):12–28.
- Baayen, R. H., Milin, P., Durevodic, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological review*, 118(3):438–481.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). The celex lexical database (cd-rom).
- Bailey, T. M. and Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory and Language*, 44(4):568–591.
- Balling, L. W. and Baayen, R. H. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, 125(1):80–106.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, pages 445–459.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.
- Berent, I., Steriade, D., Lennertz, T., and Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104(3):591–630.
- Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., and Grainger, J. (2016).

- Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review*, 23(2):533–539.
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98.
- Chuang, Y.-Y., Voller, M.-l., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., and Baayen, R. H. (2019). On the processing of nonwords in word naming and auditory lexical decision. In *Proceedings of the 19th International Congress of Phonetic Sciences*, pages 1–5, Melbourne, Australia.
- Chuang, Y.-Y., Voller, M.-l., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., and Baayen, R. H. (2020). The processing of pseudoword form and meaning in production and comprehension: A computational modeling approach using Linear Discriminative Learning. preprint, PsyArXiv.
- Coleman, J. and Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology*.
- Connine, C. M., Blasko, D. G., and Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, 30(2):234–250.
- Cook, N. R., Gillman, M. W., Rosner, B. A., Taylor, J. O., and Hennekens, C. H. (1997). Prediction of young adult blood pressure from childhood blood pressure, height, and weight. *Journal of Clinical Epidemiology*, 50(5):571–579.
- Dahan, D., Magnuson, J. S., and Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive psychology*, 42(4):317–367.
- Davies, M. (2008). The Corpus of Contemporary American English (COCA).
- Diaz, M. T., Karimi, H., Troutman, S. B. W., Gertel, V. H., Cosgrove, A. L., and Zhang, H. (2021). Neural sensitivity to phonological characteristics is stable across the lifespan. *NeuroImage*, 225:117511.
- Dupoux, E. and Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, 29(3):316–335.
- Emmorey, K. D. (1989). Auditory morphological priming in the lexicon. *Language and Cognitive Processes*, 4(2):73–92.
- Ernestus, M. and Cutler, A. (2015). Baldey: A database of auditory lexical decisions. *The*

- Quarterly Journal of Experimental Psychology*, 68(8):1469–1488.
- Feldman, L. B., O'Connor, P. A., and del Prado Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 16(4):684–691.
- Ford, C., Nenadić, F., Brenner, D., and Tucker, B. V. (2018). Shorter phone duration facilitates isolated spoken word recognition. In *Proceedings of the 11th International Conference on the Mental Lexicon*, Edmonton, Alberta, Canada.
- Frisch, S. A., Large, N. R., and Pisoni, D. B. (2000). Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords. *Journal of memory and language*, 42(4):481–496.
- Gierut, J. A., Morrisette, M. L., and Ziemer, S. M. (2010). Nonwords and Generalization in Children With Phonological Disorders. *American Journal of Speech-Language Pathology*, 19(2):167–177. Publisher: American Speech-Language-Hearing Association.
- Glanzer, M. and Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 2(1):21–31. Place: US Publisher: American Psychological Association.
- Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of memory and language*, 28(5):501–518.
- Goodman, E. and Bates, J. C. (1997). On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. *Language and Cognitive Processes*, 12(5-6):507–584.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4):267–283.
- Hendrix, P. and Sun, C. C. (2020). A Word or Two About Nonwords: Frequency, Semantic Neighborhood Density, and Orthography-to-Semantics Consistency Effects for Nonwords in the Lexical Decision Task. *Journal of Experimental Psychology*.
- Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology*, 48(2):106–112. Place: US Publisher: American Psychological Association.
- Hulden, M. (2009). Foma: a finite-state compiler and library. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages

- 29–32. Association for Computational Linguistics.
- Janse, E. (2009). Neighbourhood density effects in auditory non-word processing in aphasic listeners. *Clinical linguistics & phonetics*, 23(3):196–207.
- Janse, E. and Newman, R. S. (2013). Identifying nonwords: Effects of lexical neighborhoods, phonotactic probability, and listener characteristics. *Language and Speech*, 56(4):421–441.
- Järvikivi, J., Pykkönen, P., and Niemi, J. (2009). Exploiting degrees of inflectional ambiguity: Stem form and the time course of morphological processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(1):221–237.
- Järvikivi, J. and Pykkönen, P. (2011). Sub- and Supralexical Information in Early Phases of Lexical Access. *Frontiers in Psychology*, 2.
- Kapatsinski, V. (2005). Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. Progress Report 27, Indiana University Speech Research Laboratory, Bloomington, IN.
- Kelley, M. C. (2020). Phonetics.jl.
- Kelley, M. C. and Tucker, B. V. (2022). Using acoustic distance and acoustic absement to quantify lexical competition. *The Journal of the Acoustical Society of America*, 151(2):1367–1379.
- Keuleers, E. and Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, 42(3):627–633.
- Kotzor, S., Wetterlin, A., and Lahiri, A. (2017). Symmetry or asymmetry: Evidence for underspecification in the mental lexicon. In *The Speech Processing Lexicon: Neurocognitive and Behavioural Approaches*, pages 85–106. De Gruyter Mouton.
- Kučera, H. and Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.
- Lahiri, A. (2018). Predicting universal phonological contrasts. In *Phonological Typology*, pages 229–272. De Gruyter Mouton.
- Lahiri, A. and Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38(1):44–59.
- Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. research on speech perception. technical report no. 6. Technical report, Department of Psychology, Indiana University, Bloomington, IN.
- Luce, P. A., Goldinger, S. D., Auer, E. T., and Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62(3):615–625.

- Luce, P. A. and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, 19(1):1.
- Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In Altmann, G. T. M., editor, *Cognitive models of speech processing*, pages 122–147. MIT Press, Cambridge, MA.
- Lázaro, M., Illera, V., and Sainz, J. (2016). The suffix priming effect: Further evidence for an early morpho-orthographic segmentation process independent of its semantic content. *The Quarterly Journal of Experimental Psychology*, 69(1):197–208.
- Lõo, K. and Järviķivi, J. (2019). Whole-word frequency effects in English masked priming: Very little CORN in CORNER and CORNET. In *Proceedings of The 11th International Conference on the Mental Lexicon*, pages 1–4.
- Magnuson, J. S., You, H., Luthra, S., Li, M., Nam, H., Escabí, M., Brown, K., Allopenna, P. D., Theodore, R. M., Monto, N., and Rueckl, J. G. (2020). EARSHOT: A Minimal Neural Network Model of Incremental Human Speech Recognition. *Cognitive Science*, 44(4):e12823.
_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.12823>.
- Marslen-Wilson, W. and Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8(1):1–71.
- Marslen-Wilson, W. and Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, 15(3):576.
- Marslen-Wilson, W. D. and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1):29–63.
- McClelland, J. L. and Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1):1–86.
- Morris, J., Porter, J. H., Grainger, J., and Holcomb, P. J. (2011). Effects of lexical status and morphological complexity in masked priming: An ERP study. *Language and Cognitive Processes*, 26(4-6):558–599.
- Moscoso del Prado Martiń, F., Kostic, A., and Baayen, R. H. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition*, 94(1):1–18.
- Nelson, N. R. and Wedel, A. (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics*, 64(Supplement C):51–70.
- Nenadić, F., ten Bosch, L., and Tucker, B. V. (2018). Implementing DIANA to model isolated

- auditory word recognition in English. In *Proceedings of Interspeech 2018*, pages 3772–3776, Hyderabad.
- Nenadić, F. and Tucker, B. V. (2020). Computational modelling of an auditory lexical decision experiment using jTRACE and TISK. *Language, Cognition and Neuroscience*, 0(0):1–29. Publisher: Routledge .eprint: <https://doi.org/10.1080/23273798.2020.1764600>.
- Norris, D. and McQueen, J. M. (2008). Shortlist b: A bayesian model of continuous speech recognition. *Psychological Review*, 115(2):357–395.
- Norris, D., McQueen, J. M., and Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23:299–325. Publisher: Cambridge University Press.
- Öhman, S. E. G. (1966). Coarticulation in VCV Utterances: Spectrographic Measurements. *The Journal of the Acoustical Society of America*, 39(1):151–168.
- Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In Bod, R., Hay, J., and Jannedy, S., editors, *Probabilistic Linguistics*, pages 177–228. MIT Press.
- Pitt, M. A., Dille, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). Buckeye corpus of conversational speech (2nd release). *Columbus, OH: Department of Psychology, Ohio State University*.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother’s brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11(6):1090–1098.
- Revill, K. P., Aslin, R. N., Tanenhaus, M. K., and Bavelier, D. (2008). Neural correlates of partial lexical activation. *Proceedings of the National Academy of Sciences*, 105(35):13111–13115. Publisher: National Academy of Sciences Section: Social Sciences.
- Richtsmeier, P. T. (2011). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology*, 2(1):157–183. Publisher: De Gruyter Mouton Section: Laboratory Phonology.
- Scharenborg, O. (2010). Modeling the use of durational information in human spoken-word recognition. *The Journal of the Acoustical Society of America*, 127(6):3758–3770. Publisher: Acoustical Society of America.
- Shafaei-Bajestan, E., Moradipour-Tari, M., Uhrig, P., and Baayen, R. H. (2020). LDL-AURIS: Error-driven Learning in Modeling Spoken Word Recognition. Technical report, PsyArXiv.

- Stovitz, S. D., Pereira, M. A., Vazquez, G., Lytle, L. A., and Himes, J. H. (2008). The Interaction of Childhood Height and Childhood BMI in the Prediction of Young Adult BMI. *Obesity*, 16(10):2336–2341. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1038/oby.2008.359>.
- ten Bosch, L., Boves, L., and Ernestus, M. (2013). Towards an end-to-end computational model of speech comprehension: Simulating a lexical decision task. In *INTERSPEECH-2013*, pages 2822–2826, Lyon.
- ten Bosch, L., Boves, L., and Ernestus, M. (2015a). Diana, an end-to-end computational model of human word comprehension. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, pages 1–4, Glasgow.
- ten Bosch, L., Boves, L., Tucker, B. V., and Ernestus, M. (2015b). DIANA: Towards computational modeling reaction times in lexical decision in North American English. In *INTERSPEECH-2015*, pages 1576–1580, Dresden.
- ten Bosch, L., Ernestus, M., and Boves, L. (2018). Analyzing Reaction Time Sequences from Human Participants in Auditory Experiments. In *Interspeech 2018*, pages 971–975. ISCA.
- Teruya, H. and Kapatsinski, V. (2019). Deciding to look: revisiting the linking hypothesis for spoken word recognition in the visual world. *Language, Cognition and Neuroscience*, 34(7):861–880. Publisher: Routledge eprint: <https://doi.org/10.1080/23273798.2019.1588338>.
- Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71:249–267.
- Tsang, Y.-K. and Chen, H.-C. (2013). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language*, 68(3):223–239.
- Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., and Sims, M. (2019). The Massive Auditory Lexical Decision (MALD) database. *Behavior Research Methods*, 51(3):1187–1204.
- Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2):270–278. Place: US Publisher: American Psychological Association.
- Vitevitch, M. S., Armbruster, J., and Chu, S. (2004). Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):514–529.

- Vitevitch, M. S. and Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science (0956-7976)*, 9(4):325–329.
- Vitevitch, M. S. and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in english. *Behavior Research Methods, Instruments, & Computers*, 36(3):481–487.
- Vitevitch, M. S. and Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics*, 2:75–94.
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., and Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech*, 40(1):47–62.
- Vitevitch, M. S. and Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, 3(1):64–73.
- Vroomen, J. and de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3):710–720.
- Weide, R. (2005). The Carnegie Mellon pronouncing dictionary [cmudict. 0.6]. *Pittsburgh, PA: Carnegie Mellon University*.
- Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & psychophysics*, 35(1):49–64. Publisher: Springer.
- Xiang, M., Harizanov, B., Polinsky, M., and Kravtchenko, E. (2011). Processing morphological ambiguity: An experimental investigation of Russian numerical phrases. *Lingua*, 121(3):548–560.
- Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., and Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English Lexicon Project. *Journal of experimental psychology. Learning, memory, and cognition*, 41(3):597–613.
- Yates, M., Locker, L., and Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, 11(3):452–457.
- You, H. and Magnuson, J. S. (2018). TISK 1.0: An easy-to-use Python implementation of the time-invariant string kernel model of spoken word recognition. *Behavior Research Methods*, 50(3):871–889.