Check for updates

Consistency norms for 37,677 english words

Qian Wen Chee [1] · Keng Ji Chow [1] · Melvin J. Yap [1] · Winston D. Goh [1]

Abstract

Consistency reflects the mapping between spelling and sound. That is, a word is feedforward consistent if its pronunciation matches that of similarly spelled words, and feedback consistent if its spelling matches that of similar pronounced words. For a quasi-regular language such as English, the study of consistency effects on lexical processing has been limited by the lack of readily accessible norms. In order to improve current methodological resources, feedforward (spelling-to-sound) and feedback (sound-to-spelling) consistency measures for 37,677 English words were computed. The consistency measures developed here are operationalized at the composite level for multisyllabic words, and at different sub-syllabic segments (onset, nucleus, coda, oncleus, and rime) for both monosyllabic and multisyllabic words. These measures constitute the largest database of English consistency norms to be developed, and will be a valuable resource for researchers to explore the effects of consistency on lexical processes, such as word recognition and spelling. The norms are available as supplementary material with this paper.

Keywords Consistency · Feedforward · Feedback · Word norms · Word recognition

The relationship between spelling and sound is a key factor in determining skilled reading, and has been the focus of a number of influential models of word recognition (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007; Plaut, McClelland, Seidenberg, & Patterson, 1996). In English, however, the study of spelling-to-sound relations is complicated by much ambiguity in the way letters map onto sounds. As neatly illustrated by the poem "Our Strange Lingo", a given letter is often pronounced differently in various words:

> Beard is not the same as heard,
> Cord is different from word.
> Cow is cow, but low is low,
> Shoe is never rhymed with foe!

An important implication of these irregularities in letter-to-sound correspondences is that the interface between orthography and phonology is a fundamental issue in understanding how words are processed in English. Specifically, how are we able to resolve ambiguities in spelling and pronunciation quickly and accurately enough to read and write? Researchers have therefore sought to examine how lexical processes are influenced by spelling-to-sound relationships, that is, the *consistency* with which the phonology of a given word can be derived from its spelling, and vice versa.

Defining consistency: Feedforward vs. feedback measures

Traditionally, consistency has been defined as the mapping between spelling and sound; a word is consistent if its pronunciation matches that of similarly spelled words (Glushko, 1979). For instance, "pint" is inconsistent because its pronunciation conflicts with that of similarly spelled words, such as "mint", "hint", and "lint", whereas "face" is consistent because its pronunciation matches that of "race", "lace", and "pace".

Subsequently, Stone, Vanhoy and Van Orden (1997) put forth an intriguing proposition that it is meaningful to also consider the converse mapping between sound and spelling. According to the recurrent network theory of word perception

✉ Qian Wen Chee
psycqw@nus.edu.sg

[1] Department of Psychology, National University of Singapore, Singapore 117570, Singapore

(Stone & Van Orden, 1994), they argued that there is a bidirectional flow of activation between orthography and phonology. That is, when a pattern of orthographic codes activates a series of phonological units, the phonological units, in turn, feed activation back to the orthographic codes (Van Orden & Goldinger, 1994). Therefore, exploring inconsistencies in both directions is critical to understanding lexical processing. By this definition, the relationship between spelling and sound is henceforth described as *feedforward* consistency, whilst the mapping between sound and spelling is termed *feedback* consistency (see Fig. 1). Mapping directions are defined similarly regardless of modality; that is, feedforward consistency is always spelling-to-sound consistency, and feedback consistency is always sound-to-spelling consistency, regardless of whether stimulus input is visual or auditory.

Feedforward and feedback measures of consistency are dissociable. For instance, "roar" is feedforward consistent because all other words ending with '-oar' (e.g., "soar", "boar", "hoar") are pronounced with /ɔr/, but "roar" is also feedback *in*consistent because several other words pronounced with /ɔr/ (e.g., "core", "more", "bore") are spelled with '-ore' instead. As such, feedforward and feedback measures differ on the direction of mappings.

Calculating consistency

Pioneer work in this field initially treated consistency as a binary measure (Glushko, 1979); words were considered consistent only if all other words with similar spellings shared the same pronunciation (hereafter referred to as 'friends'), and were categorized as inconsistent so long as any other similarly spelled word had a different pronunciation (hereafter referred to as 'enemies') (Jared, McRae, & Seidenberg, 1990; Jared & Seidenberg, 1990; Ziegler, Stone, & Jacobs, 1997). However, this view ignored the *degree* of consistency—some inconsistent words had very few enemies, whereas others had many. According to contemporary perspectives, consistency is now best described as a continuous variable with graded effects (Jared et al., 1990; Treiman, Mullennix, Bijeljac-Babic, &

Richmond-Welty, 1995). Specifically, *type consistency* is calculated by dividing a given word's number of friends by its total number of friends and enemies.

$$Type\ consistency = \frac{no.of\ friends}{no.of\ friends + enemies}$$

Another measure of consistency is based on *token* estimates, which weights a given word's friends and enemies by frequency of occurrence (Treiman et al., 1995). Specifically, token consistency is computed by dividing a given word's log frequencies of friends by its total log frequencies of friends and enemies.

$$Token\ consistency = \frac{\sum logfreq\ (friends)}{\sum logfreq\ (friends) + \sum logfreq\ (enemies)}$$

For both type and token consistency, the resulting consistency ratio ranges from 0 (very inconsistent) to 1 (highly consistent).

Consistency at various grain sizes

Consistency can also be operationalized at different sub-syllabic grain sizes. By definition, a syllable comprises various sub-syllabic constituents–the onset, nucleus, coda, rime (Vennemann, 1988), and oncleus. The vowel of a syllable is always the nucleus, and the onset of a syllable comprises any consonants that precede the vowel, while the consonants that follow the vowel constitutes the coda. The concatenation of the vowel and the coda forms a higher-order grouping known as rime, which represents the word body, and the concatenation of the onset and the nucleus is the oncleus. For instance, the word "flash" (transcribed /flæʃ/), would have /fl/ as onset, /æ/ as nucleus, and /ʃ/ as coda. /flæ/ is therefore the oncleus, and /æʃ/ is the rime.

Most researchers have manipulated consistency only at the level of the rime for monosyllabic words (e.g., Jared, 1997; Jared et al., 1990; Lacruz & Folk, 2004; Peereman, Content, Bonin, 1998), because the rime is a salient unit in reading (De Cara & Goswami, 2002; Treiman & Kessler, 1995; Ziegler &
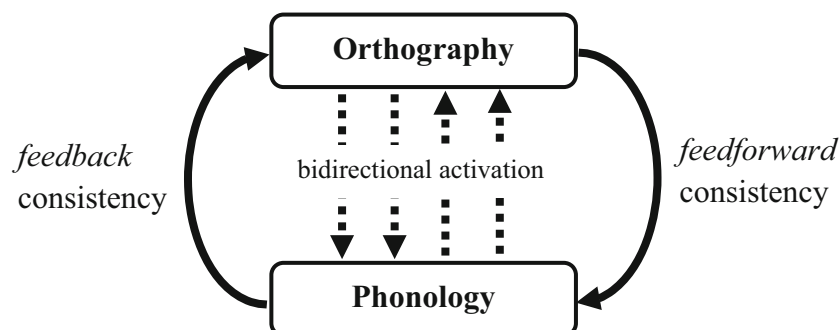

Fig. 1 Model of the bidirectional flow of activation between orthography (spelling) and phonology (sound). Feedforward consistency is the mapping from orthography to phonology, whereas feedback consistency is the mapping from phonology to orthography

Goswami, 2005), and consonant onsets vary less in their spellings and pronunciations (Kessler & Treiman, 2001). For multisyllabic words, studies have also attempted to target rime consistency by focusing on an inconsistent rime in either syllable for disyllabic words (Jared & Seidenberg, 1990), or averaging rime consistency across syllables for longer words (Yap, 2007; Yap & Balota, 2009). Nonetheless, this does not rule out the possibility that consistency measured at other sub-syllabic segments can have effects on lexical processing. For example, some studies have found that onset consistency is also a significant predictor of word recognition (Balota et al., 2004; Treiman et al., 1995; Yap & Balota, 2009), adding support to the theory that sub-syllabic representations are usually contained by an onset-rime structure in English (Kessler & Treiman, 1997; Treiman & Kessler, 1995).

Effects of consistency on lexical processing

Speeded naming and visual word recognition Empirical support for the effects of consistency on lexical processing has mostly come from findings that consistency influences performance in naming and visual lexical decision tasks. The most robust finding is that feedforward rime consistency significantly facilitates response latencies and accuracy rates (Balota et al., 2004; Chateau & Jared, 2003; Jared et al., 1990; Jared & Seidenberg, 1990; Treiman et al., 1995; Yap & Balota, 2009; Ziegler et al., 1997b; Ziegler, Petrova, & Ferrand, 2008); words with highly consistent rimes are generally responded to faster and with higher accuracy.

In contrast to the stability of effects found for feedforward consistency, however, substantial disagreement remains regarding the effects of feedback consistency on visual word recognition. Early claims of feedback consistency effects (Stone et al., 1997; Ziegler, Montant, & Jacobs, 1997) were not supported by later studies (Kessler, Treiman, & Mullennix, 2008; Peereman et al., 1998; Ziegler et al., 2008), thereby challenging the theoretical implications of Stone et al.'s (1997) hypothesis and leaving the existence of feedback consistency effects unclear. Researchers have sought to explain the apparent heterogeneous findings in the literature in terms of differences in experimental methodology, such as the failure to sufficiently control for other lexical covariates known to influence word recognition, including subjective familiarity, word frequency, and neighborhood densities (Lacruz & Folk, 2004; Massaro & Jesse, 2005; Peereman et al., 1998).

Spoken word recognition Although it has been difficult to obtain feedback consistency effects in the *visual* modality, there is an extensive literature reporting stable feedback effects across a wide range of auditory tasks, including lexical decision (Petrova, Gaskell, & Ferrand, 2011; Ziegler et al., 2004, 2008), shadowing (Pattamadilok, Morais, Ventura, &

Kolinsky, 2007; Ventura, Morais, Pattamadilok, & Kolinsky, 2004), rime detection (Petrova et al., 2011; Ziegler et al., 2004), and semantic and gender categorization (Peereman, Dufour, & Burt, 2009). These studies invariably show that feedback inconsistent words yield longer and more error-prone responses (Dich, 2011, 2014; Perre, Pattamadilok, Montant, & Ziegler, 2009; Ventura et al., 2004; Ziegler & Ferrand, 1998; Ziegler et al., 2004, 2008) than consistent words, and convincingly demonstrate that the orthographic representation of a word has substantial influence on the processing of its spoken form.

However, most of the studies investigating the effects of consistency on spoken word recognition have been conducted in either Portuguese (Ventura et al., 2004; Ventura, Kolinsky, Pattamadilok, & Morais, 2008; Ventura, Morais, & Kolinsky, 2007), or French (Pattamadilok et al., 2007; Peereman et al., 2009; Perre & Ziegler, 2008; Petrova et al., 2011; Ziegler et al., 2004), and it is not obvious whether these findings can be generalized to English. As Katz and Frost (1992) described in their concept of orthographic depth, the transparency of the relation between spelling and sound varies widely between orthographies. Given that orthographic depth is a primary source of cross-linguistic differences in word recognition (Ziegler et al., 2010), it remains to be seen whether consistency effects are language-dependent, or stable across languages.

Spelling Consistency effects have perhaps been least widely explored in spelling research. Spelling from dictation concerns the decoding of specific orthographic codes that can be assigned to a string of phonological units, and one might expect feedback consistency to play an essential role in understanding differences in spelling performance.

Most research in this domain has focused on the effects of consistency in developing readers (e.g., Alegria & Mousty, 1996; Beyermann & Penke, 2014; Caravolas, Kessler, Hulme, & Snowling, 2005; Davies & Weekes, 2005; Lété, Peereman, & Fayol, 2008; Weekes, Castles, & Davies, 2006). Indeed, studies typically observe that feedback associations influence spelling accuracy amongst developing readers, such that more spelling errors are made on feedback inconsistent words than on consistent ones (e.g., Lété et al., 2008; Weekes et al., 2006). Few observations have been made, however, regarding the influence of feedforward consistency on spelling performance, and whether skilled readers would also demonstrate the same consistency effects on spelling typically found in developing readers.

Limitations of extant literature

Evidently, there is substantial interest in the effects of consistency on lexical processing across a variety of tasks. However, while the studies cited above constitute an important step

towards understanding the influence of spelling-sound relationships on word recognition, interpretation of the current literature is hindered by several limitations.

Experimental confounds As aforementioned, the problem of confounds pervades research on consistency effects. Perhaps the most pernicious reason is because, to date, apart from a few studies (Balota et al., 2004; Chateau & Jared, 2003; Kessler et al., 2008; Treiman et al., 1995; Yap & Balota, 2009), all studies examining consistency effects have been based on factorial designs. Given the multiple variables that have been identified to influence lexical processing, it is challenging to design a factorial study that perfectly equates lexical stimuli on all variables besides consistency. As such, previous studies may not have adequately controlled for extraneous variables. Additionally, beyond the obvious differences in the range of variables controlled by separate studies, disparate stimuli selection across studies will also preclude cross-study inferences on the consistency effects in lexical processing.

Monosyllabic vs. multisyllabic words Aside from limitations due to experimental design, a significant limitation of previous studies is that they have been largely concerned with understanding consistency effects in monosyllabic words, with only a few studies extending their research to subsets of multisyllabic words (e.g., Chateau & Jared, 2003; Jared & Seidenberg, 1990; Yap & Balota, 2009). The disparity between monosyllabic and multisyllabic research on consistency is unsurprising, given the contrasting approaches to defining what constitutes consistency across multiple syllables, as previously discussed. This is troubling because most words in English have more than one syllable; as reference, almost 85% of the words in the restricted English Lexicon Project (ELP; Balota et al., 2007) are multisyllabic. Considering that reading multisyllabic words is generally a more sophisticated process (Perry, Ziegler, & Zorzi, 2010), it remains to be seen whether consistency effects exert the same influence on the processing of multisyllabic words as compared to monosyllabic words. Drawing conclusions about the effects of consistency at this stage from the current body of studies is therefore problematic because interpretation of extant findings may not be generalizable to multisyllabic words.

Measures of consistency Another area of concern is that almost all studies have manipulated consistency only for word bodies, and it is far from obvious whether consistency effects should only manifest themselves at the level of the rime and nowhere else. For instance, research has shown that the pronunciation of vowels can vary systematically with the identity of the preceding consonant (Treiman, Kessler, & Bick, 2003; Treiman, Kessler, Zevin, Bick, & Davis, 2006), and therefore, whilst the rime is especially salient in word recognition, it is

most likely not the only unit relevant to pronunciation (Jared et al., 1990).

Additionally, researchers have generally focused exclusively on either feedforward or feedback consistency, and it is unclear whether the type of consistency effects observed are limited by the modality of stimulus input. Findings on the effects of feedback consistency on visual word recognition have been mixed, and to date, there is no comparative study on the influence of feedforward consistency on spelling and spoken word recognition, making it difficult to reconcile results with the broader theory of recurrent networks in word perception (Stone & Van Orden, 1994; Stone et al., 1997; Van Orden & Goldinger, 1994).

Aims of the study

In order to overcome the above limitations, it is crucial for both feedforward and feedback consistency effects to be examined in depth and at various grain sizes, without methodological issues that can potentially confound results. One way to do so is by means of a large-scale study analyzed using regression methods. This megastudy advantage allows for the exploration of lexical properties that influence word recognition without any constraints (see Balota et al., 2004). Whilst five studies have already used this approach to examine consistency effects, they were restricted to either monosyllabic words (Balota et al., 2004; Kessler et al., 2008; Treiman et al., 1995), or monomorphemic words (Chateau & Jared, 2003; Yap & Balota, 2009).

To improve the generalizability of their results, a more comprehensive approach would be to examine consistency effects at different levels for a range of word lengths. However, this can only be done using a sizeable and reliable database of consistency norms. Unfortunately, such a database is currently lacking in the field. Ziegler et al. (1997b) only computed norms for the rime segments of 2694 monosyllabic, monomorphemic words, and Yap and Balota (2009) computed norms for the onset and rime segments of 9643 monomorphemic words based on the ELP transcriptions, some of which were erroneous. Additionally, Yap and Balota's (2009) dataset is limited to type consistency measures for the first syllable, or averaged across syllables, and due to an error, all first syllable feedforward and feedback onset consistency values were reduced to binary 0/1 values.

The aim of this study was therefore twofold: firstly, to create and validate a large set of feedforward and feedback consistency norms, operationalized at different subsyllabic segments, for both monosyllabic and multisyllabic words, and secondly, to explore the effects of consistency in detail by entering the norms as predictors of behavioral data in a series of regression models on speeded pronunciation, and both visual and auditory lexical decision performance.

Method

This section illustrates how stimuli were selected, and how the consistency norms were eventually computed after a process of tokenization, syllabification, phonological parsing, and orthographic parsing. Other than some manual adjustments, all functions were performed almost fully automatically using Python 3.6.3 (Python Software Foundation, n.d.).

Stimuli

The words in this database were retrieved from the ELP (Balota et al., 2007), because it is by far the largest available corpus of English that provides standardized behavioral data for a plethora of monosyllabic and multisyllabic words. From the restricted lexicon, 40,481 words were obtained. A review of the transcriptions provided in the ELP revealed several inaccuracies (e.g., *knobbly* (ELP SAMPA: n"Ab.bl%aI ; IPA: /ˈnɑblaɪ/ instead of /ˈnɑbli/), and *lipid* (ELP SAMPA: "laIpId ; IPA: /ˈlaɪpɪd/ instead of /ˈlɪpɪd/), etc.), presumably because the ELP transcriptions were compiled from different sources by a single linguist. Therefore, the American English pronunciations of the words were instead retrieved from Oxford Dictionaries (Oxford University Press, 2020), which is regularly maintained by a skilled team of lexicographers, via application programming interface (API).

Of the 40,481 words in the ELP, 26,604 transcriptions were returned from Oxford Dictionaries, as pronunciations for inflections were mostly not provided. For cases where multiple pronunciations were returned from the application, the first pronunciation was selected, which reflects the most important and common entry of the word in Oxford Dictionaries' database (Oxford English Dictionary, 2020). The remaining 13,877 words were lemmatized, and then, wherever possible, manually transcribed based on compound rules for word formation. For example, to represent the regular past inflection '–ed', /t/, /d/, or /əd/ was appended to the end of the transcribed lemma form, depending on the voicing of the final phoneme. Based on this method, a further 11,073 words were successfully transcribed, resulting in a final set of 37,677 words and their corresponding pronunciations. Of the remaining 2804 words from the ELP that were not transcribed, 1,661 words are either proper nouns (e.g., "Abigail", "Adele", etc.), enclitic groups (e.g., "how'd", "that'll", "kid's", etc.), or a combination of both (e.g., "Kate's", etc.), and the transcriptions of 1443 words could not be obtained even in lemma form (e.g., "aah", etc.). All phonetic transcriptions follow the conventions of the International Phonetic Association.

Schwa insertion and elision For the purpose of mappings, a couple of modifications were made to schwa /ə/ representations in the transcriptions. In English, the schwa is a neutral vowel, and its precise realization varies depending on adjacent consonants. When two consonants cannot be co-articulated, the schwa can be inserted as an epenthetic vowel to make the phonemic sequence phonotactically legal (Halle & Idsardi, 1997; Haselwood, 2007). For instance, "rhythm" is typically pronounced as /ˈrɪð(ə)m/, instead of /ˈrɪðm/, because the consonant /m/ cannot be co-articulated with the preceding consonant /ð/.

On the other hand, schwas may also exist in the phonological representations of some words, but are not realized during speech (van Oostendorp, 1995), due to a phonological variation process known as schwa elision (Dalby, 1986). Schwa elision is typically conditioned by linguistic and extralinguistic factors, such as lexical stress position (Hooper, 1978), and speech style and tempo (Zwicky, 1972). Several researchers also suggest that phonotactics play a role in schwa deletion, such that the deletion is likely to occur in environments where the two consonants straddling the schwa can legally combine to form an onset cluster (Hammond, 1999; Patterson, LoCasto, & Connine, 2003).

The treatment of schwas by Oxford Dictionaries is somewhat arbitrary. Some transcriptions practice an extreme approach, and omit schwas when their occurrence can be predicted by consonants that cannot be co-articulated. For example, "flatten" is simply transcribed as /ˈflætn/, without a schwa between /t/ and /n/, even though /tn/ is a phonotactically illegal within-syllable sequence. Other transcriptions represent schwas in parentheses to indicate an indistinguishable utterance, even when the schwa is unlikely to be realized in speech. For instance, "baffling" is transcribed /ˈbæf(ə)lɪŋ/, with a schwa inserted between /f/ and /l/, even though the vowel is unlikely to be pronounced.

This unpredictable representation of schwas is problematic for mappings between orthography and phonology, because it is unclear when schwas are represented in writing or should be a feature only in speech. Hence, two manual corrections were made in order to better capture the realization of schwas at speech output level. Firstly, schwas that were present in transcriptions but were not likely to be pronounced were manually deleted. For example, the obtained transcription for "scrambling" was /ˈskræmb(ə)lɪŋ/, but the schwa was deleted to give /ˈskræmblɪŋ/ as the final transcription. Secondly, schwas were manually inserted between phonotactically illegal consonants, if they occurred within the same syllable. For instance, a schwa was inserted between /t/ and /n/ for the transcription of "kitten" (/ˈkɪtn/), such that the final transcription was /ˈkɪt(ə)n/. Care was taken to ensure that all resulting transcriptions were phonotactically legal.

Tokenization, syllabification and phonological parsing

After finalizing the phonologies of the words, the phoneme strings were organized into sub-syllabic onset, nucleus, and

coda segments. To achieve this, the transcriptions of the words were first tokenized into individual phonemes based on the Oxford Dictionaries' key to pronunciation[1], which recognizes diphthongs as single sound segments. Following this, each vowel was appended to the nucleus of a syllable, beginning from the first syllable, until all vowels had been assigned.

The assignment of intervocalic consonants was then based on the maximum onset principle (Kahn, 2015). Whilst determining syllabic boundaries has proved notoriously difficult (Jensen, 2000), this widely accepted rule-based algorithm for syllabic division states that intervocalic consonants should be maximally assigned to onset positions, unless their inclusion in onset is phonotactically illegal (Clements & Keyser, 1983). In English, a legal onset is patterned by the sonority hierarchy and aligns with the $C_1.C_2.C_3.C_4$ template; C_1 is /s/, C_2 is an element of all other obstruents and nasal stops, C_3 is from the set of liquids /l/ and /r/, and C_4 is from the set of glides /j/ and /w/.

The phonological parsing process can thus be illustrated using /ˈæbˌstrækt/ (transcription for "abstract"). The parser first assigned /æ/ to the nuclei of both the first and second syllable. Then, for intervocalic consonants /bstr/, /str/ was appended to onset of the second syllable, while /b/ was affixed as coda of the first syllable, since no consonant is permitted to precede /s/ in onset. Finally, /kt/ was assigned to coda of the second syllable.

Higher-order oncleus and rime groupings were then obtained by concatenating the onset and nucleus for the oncleus segment, and the nucleus and coda for the rime segment respectively.

Orthographic parsing

Although sub-syllabic structure is primarily defined at the phonological level, it was also necessary to parse the spelling of each word into sub-syllabic segments for the eventual calculation of consistency. Using a list of predefined phoneme-grapheme correspondences (see Appendix A), this was done by identifying possible orthographic representations at the syllabic boundaries based on the parsed phonological representations.

For example, /ˈækʃ(ə)n/ (transcription for "action") had the consonants /k/, /ʃ/ and /n/ at its syllabic boundaries after the process of phonological parsing. According to the list of correspondences, /k/ can be represented by 'c', /ʃ/ by 't' and /n/ by 'n'. Therefore, 'c' was assigned to orthographic coda of the first syllable, followed by 't' into onset of the second syllable, and 'n' in coda of the second syllable. Interconsonantal letter strings were assumed to represent vowels and were thus automatically assigned as orthographic nucleus. This approach

achieved orthographic parsing to a high degree of accuracy, although some manual corrections were required due to irregularities in the English spelling.

Manual adjustments

Before the calculation of consistency, the orthographic and phonological alignments for all words were visually inspected and corrected for minor inconsistencies. As described below, a couple of manual adjustments were made to the phonological and orthographic parsings, in order to optimize the mappings for the computation of consistency.

The grapheme 'x' Firstly, it was necessary to make the assumption that there is no ambisyllabicity at both the orthographic and phonological level—that is, a single grapheme or phoneme cannot be shared across syllables (Kahn, 2015; Jensen, 2000). This posed a problem for the grapheme 'x', which has no distinctive sound of its own and instead variably represents consonants clusters such as /ks/, /gz/, or /ŋz/. Syllabic parsing by the maximum onset principle typically entailed these consonant clusters splitting across syllabic boundaries, with the first phoneme occurring in coda of the preceding syllable, and the second phoneme initiating the onset of the following syllable. For example, the phonological parsing for "extra" was /ˈɛk.strə/, with /k/ occurring in coda of the first syllable, and /s/ appended to onset of the second syllable, even though the consonant cluster /ks/ is represented by a single grapheme 'x'.

This would have complicated the calculation of consistency for the grapheme 'x', because it would have been illogical to compute two separate onset and coda mappings for a single entity. Therefore, an additional sequencing constraint was implemented for consonant clusters /gz/ and /ks/ to always take on the same sub-syllabic affiliation. Specifically, these consonants clusters were deliberately parsed into the coda positions of the preceding syllable, since /gz/ and /ks/ occurring in onset positions of the following syllable would violate the sonority hierarchy that is a requisite of the maximum onset principle.

Reversal of liquid-vowel segments The automated orthographic parsing process was also completed under the presumption that the order of letters in spelling corresponds linearly to the order of phonemes in pronunciation. However, this was not the case for most words ending in liquid-vowel rimes, such as "table" and "timbre", in which the pronunciation of glides /l/ and /r/ follow the schwas, despite its orthographic code preceding the vowel. This posed a problem for onset and coda consistency mappings. Based on the original parsing, the orthographic onset for a suffix '-ble' was 'bl', even though the phonological onset was just /b/. The orthographic coda had no representation, even though /l/ was in phonological coda.

In order to resolve this complication, it was necessary to consider the etymology of such words, which can be traced to borrowings from Old French into Middle English. For example, many adjectives in English were formed by the addition of a French suffix '-able' (as in "drinkable" and "curable") and '-ible' (as in "edible" and "audible") to English roots (Bloomfield, 1933). Other words such as "table", "stable", and "ogre" were also borrowed directly from French terms.

In French, these words were pronounced without the second vowel (e.g., "tabl"), as it is phonotactically legal for obstruent-liquid clusters (e.g., /bl/) to be articulated in phonological coda. In order to safeguard the articulatory orientation of English's prevailing sound system, then, after these words were borrowed from French, it is hypothesized that a schwa was inserted between the phonotactically illegal clusters as part of an epenthetic process (Halle & Idsardi, 1997). Since the word's spelling was still retained, this inadvertently resulted in a reverse order of liquid-vowel phonemes vis-à-vis its spelling.

In order to homogenize the correspondence between the orthographic code and its pronunciation, therefore, all syllables spelled with a reversed liquid-vowel order had their orthographic parsings deliberately re-aligned such that the liquid (typically 'r' and 'l') was removed from onset and appended to coda. In forming the orthographic rime, nonetheless, order of spelling was retained. To illustrate this, the final alignment for the suffix '-ble' had 'b' in onset, 'e' in nucleus, 'l' in coda, 'be' as oncleus, and 'le' as rime. While this re-organization may seem unnatural, this diachronic approach sufficiently allowed for a more balanced mapping between orthography and phonology.

Computing the norms

Once the orthographic and phonological alignments for all words in the database were finalized, the final step was to compute the feedforward and feedback consistency measures for individual sub-syllabic segments for each syllable. As described in the introduction, consistency was calculated based on both type and token estimates. Consistency was also computed in a position-specific manner, because Yap (2007) demonstrated that position-specific computations outperformed consistency measures computed across syllables in terms of the amount of variance accounted for in naming and lexical decision tasks. Therefore, consideration of friends and enemies were based on occurrence within the same syllable. For example, feedforward rime consistency for "boot" was based only on instances of '-oot' occurring in first syllable.

Besides individual sub-syllabic consistency, composite consistency measures were also computed, with the aim of reflecting average consistency across syllabic positions (Yap, 2007; Yap & Balota, 2009). For example, the composite feedforward onset consistency of the disyllabic word "window" is the mean feedforward onset consistency of 'win' and 'dow'. Composite consistency measures were computed only for multisyllabic words, since monosyllabic words only comprise a single syllable.

Results and discussion

Type consistency norms for the monomorphemic words in this dataset were first computed and compared with the measures obtained by Yap and Balota (2009) to establish their validity. Pearson correlation coefficients between this subset of monomorphemic norms and the type consistency measures for monomorphemic words in Yap and Balota (2009) are reported in Table 1. All correlations, except for one, were above 0.750, and were highly significant (all $ps < 0.001$) for first syllable and composite onset and rime consistency measures. Overall, the correlations indicate a high degree of similarity between this subset of monomorphemic norms and those developed by Yap and Balota (2009), despite the errors in the latter set of norms and the difference in transcriptions.

Following the above correlational analyses, hierarchical regression analyses were also conducted to examine how much variance the full set of consistency norms could account for in word recognition performance, above and beyond typical variables known to influence lexical processing. The analyses were performed on the visual lexical decision data from the English Crowdsourcing Project (ECP; Mandera, Keuleers, & Brysbaert, in press), the speeded pronunciation and visual lexical decision data from the ELP (Balota et al., 2007), and on the American dialect-level auditory lexical decision data from the Auditory English Lexicon Project (AELP; Goh, Yap & Chee, in press). Similar to Yap and Balota (2009), the rationale of the hierarchy was to first enter variables coding onset and prosodic properties designed to capture variance associated with voice key biases (Balota et al., 2004; Kessler, Treiman, & Mullenix, 2002; Spieler & Balota, 1997; Treiman et al., 1995) and the stress pattern of a word (see Arciuli & Cupples, 2006; Chateau & Jared, 2003; Yap & Balota, 2009), followed by standard lexical variables that have traditionally been established to predict variance in word recognition latencies and accuracy. First syllable and composite consistency measures were entered last to assess their effects only after surface and lexical measures had been controlled for. Only onset and rime consistency measures were considered at this stage, to allow for a fair comparison with established findings in the literature.

The hierarchy of the regression for speeded pronunciation and visual lexical decision was thus as follows: Phonological onsets were entered in Step 1, and stress typicality[2] was

---

[2] Stress typicality was defined as the proportion of words that share the same stress assignment as the given word, amongst all the words that have the same number of syllables as the given word.

Table 1. Correlations with Yap & Balota ([2009](#)) consistency measures for mono-morphemic monosyllabic and multisyllabic words

Measure	Monosyllabic words ($n = 3376$)	Multisyllabic words ($n = 5636$)
Syllable 1 Feedforward onset consistency	0.827***	0.827***
Syllable 1 Feedforward rime consistency	0.893***	0.815***
Syllable 1 Feedback onset consistency	0.906***	0.919***
Syllable 1 Feedback rime consistency	0.941***	0.694***
Composite Feedforward onset consistency	NA	0.788***
Composite Feedforward rime consistency	NA	0.754***
Composite Feedback onset consistency	NA	0.863***
Composite Feedback rime consistency	NA	0.751***

***$p < .001$

entered in Step 2. All other standard lexical variables were entered in Step 3, including number of syllables, word length, number of morphemes, log-transformed HAL word frequency (Burgess & Livesay, [1998](#))[3], orthographic and phonological neighborhood densities, uniqueness points, spreads, Levenshtein distances, and Levenshtein neighborhood frequencies. Then, because some words have different spellings for American and British English (e.g., "color" and "colour"), and some words are homophonous (e.g., "for", "fore", and "four"), two binary variables coding for such instances were entered in Step 4 to control for any potential influence of ambiguities in spellings and pronunciations. First syllable onset and rime consistency measures were then entered in Step 5, followed by composite onset and rime consistency measures in Step 6. The regression for auditory lexical decision was similar, with the exception that sound file duration was included as an additional first step, and number of phonemes was entered in place of word length. The data for number of morphemes, word frequency, and orthographic and phonological Levenshtein neighborhood frequencies were taken from the ELP (Balota et al., [2007](#)). All other variables were computed based on the 37,677 items' specific orthography and phonology, and where applicable, based on these same items as base corpus.

Separate sets of analyses were conducted for monosyllabic and multisyllabic words, as stress typicality and composite consistency measures are irrelevant to words with only one syllable. Since the purpose of this study was for the creation of consistency norms, the following discussion focuses on examining the data from the final two steps (first syllable and composite consistency measures) of the regressions. Nonetheless, the results from the preceding steps were as expected.

Although analyses were completed for both the type and token norms, the data presented here are based on type

measures only, as the two sets of measures were highly correlated (rs between 0.99 and 1.00) and accounted for virtually identical amounts of variance. Tables [2](#), [3](#), [4](#), and [5](#) present the results of the regression analyses for speeded pronunciation, visual lexical decision, and auditory lexical decision performance respectively, for both z-scored reaction times (RTs) and response accuracy.

Task comparisons

First syllable and composite consistency measures accounted for the most incremental variance in response latencies for speeded pronunciation (3.3% in monosyllabic words, 2.4% in multisyllabic words), compared to visual lexical decision (ECP: 0.5% in monosyllabic words, 0.5% in multisyllabic words; ELP: 0.4% in monosyllabic words, 0.6% in multisyllabic words), and auditory lexical decision (1.0% in monosyllabic words, 0.0% in multisyllabic words). This was likewise for response accuracy; consistency accounted for more variance in speeded pronunciation (2.4% in monosyllabic words, 1.4% in multisyllabic words) than in visual lexical decision (ECP: 0.4% in monosyllabic words, 0.4% in multisyllabic words; ELP: 0.2% in monosyllabic words, 0.7% in multisyllabic words), and auditory lexical decision (1.8% in monosyllabic words, 0.3% in multisyllabic words).

Evidently, the effects of consistency are clearly task-modulated; consistency accounts for more unique variance in speeded pronunciation than in both the visual and auditory lexical decision tasks, reflecting the emphasis on phonological decoding in the pronunciation task. Since consistency affects the ease of generating articulatory programs in naming (Ziegler et al., [1997a](#), [b](#)), inconsistent words presumably take longer to name, and are more prone to response errors. Per contra, lexical decision responses are less dependent on translating between orthography and phonology (Katz et al., [2005](#), [2012](#)), and therefore it is unsurprising that spelling-to-sound correspondences are less influential in the lexical decision tasks, as compared to naming (Hino & Lupker, [2000](#)).

---

[3] Although contextual diversity (SUBTLEX-CD) is a better predictor of word processing than word frequency (Adelman, Brown, & Quesada, [2006](#); Brysbaert & New, [2009](#)), there is no SUBTLEX-CD data available for 3132 words in this study. As such, log HAL frequency was chosen for better inclusion of all words in the regression analyses.

Table 2. Standardized RT and accuracy regression coefficients from Steps 1 to 6 of the item-level regression analyses for speeded pronunciation performance for monosyllabic words and multisyllabic words from the ELP. The p value for each coefficient and R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words ($n = 5714$)		Multisyllabic words ($n = 31{,}268$)	
	RT	Accuracy	RT	Accuracy
Surface variables (onsets)				
Adjusted R^2	**0.282***	**0.000**	**0.036***	**0.003***
Surface variables (stress)				
Adjusted R^2			**0.125**	**0.018**
ΔR^2	**NA**	**NA**	**0.089***	**0.015***
Standard lexical variables				
Number of syllables	NA	NA	0.241***	-0.256***
Word length	0.371***	− 0.259***	0.182***	0.105***
Number of morphemes	− 0.086***	0.131***	− 0.086***	0.129***
Log_Freq_HAL	− 0.369***	0.406***	− 0.446***	0.405***
Orthographic N	0.120***	− 0.064†	0.019*	− 0.027**
Phonological N	− 0.001	0.034	0.010	− 0.014
Orthographic uniqueness point	− 0.125***	0.167***	− 0.045***	− 0.050***
Phonological uniqueness point	− 0.050***	0.060**	− 0.073	0.146***
Orthographic spread	− 0.077***	0.052**	− 0.039***	0.036***
Phonological spread	− 0.023†	− 0.009	0.000***	− 0.019*
OLD20	0.189***	− 0.158***	0.108***	− 0.088***
PLD20	− 0.050†	0.105**	0.096***	− 0.098***
OLDF	0.043*	− 0.106***	0.056***	− 0.041***
PLDF	0.074***	− 0.087***	0.080***	− 0.123***
Adjusted R^2	**0.531**	**0.198**	**0.534**	**0.246**
ΔR^2	**0.250***	**0.199***	**0.409***	**0.228***
Spellings and pronunciations				
Spelling (US/UK)	0.028**	− 0.038**	− 0.015***	0.010*
Homophone	− 0.021*	0.009	− 0.008*	0.000
Adjusted R^2	**0.533**	**0.199**	**0.535**	**0.246**
ΔR^2	**0.002***	**0.002**	**0.000**	**0.000**
Syllable 1 consistency variables				
Feedforward onset consistency	− 0.120***	0.085***	− 0.004	− 0.019**
Feedforward rime consistency	− 0.073***	0.129***	− 0.012*	0.024**
Feedback onset consistency	− 0.075***	0.018	− 0.054***	0.017**
Feedback rime consistency	− 0.077***	0.025†	0.002	− 0.027***
Adjusted R^2	**0.566**	**0.223**	**0.544**	**0.250**
ΔR^2	**0.033***	**0.024***	**0.009***	**0.004***
Composite consistency variables				
Feedforward onset consistency			− 0.117***	0.090***
Feedforward rime consistency			− 0.094***	0.075***
Feedback onset consistency			0.056***	− 0.034***
Feedback rime consistency			− 0.046***	0.052***
Adjusted R^2			**0.558**	**0.260**
ΔR^2	**NA**	**NA**	**0.015***	**0.010***

$*p < 0.05, **p < 0.01, ***p < 0.001, †p < 0.10$

Table 3. Standardized RT and accuracy regression coefficients from Steps 1 to 6 of the item-level regression analyses for visual lexical decision performance for monosyllabic words and multisyllabic words from the ECP. The p value for each coefficient and R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words (n = 3802)		Multisyllabic words (n = 22,462)	
	RT	Accuracy	RT	Accuracy
Surface variables (onsets)				
Adjusted R^2	**0.004****	**0.004***	**0.008*****	**0.001****
Surface variables (stress)				
Adjusted R^2			**0.054**	**0.001**
ΔR^2	NA	NA	**0.046*****	**0.001*****
Standard lexical variables				
Number of syllables	NA	NA	0.172***	− 0.098***
Word length	0.133***	− 0.087*	0.137***	0.118***
Number of morphemes	0.044**	0.045*	− 0.024***	0.099***
Log_Freq_HAL	− 0.633***	0.510***	− 0.662***	0.585***
Orthographic N	− 0.025	0.071	0.043***	− 0.021†
Phonological N	0.030	− 0.005	0.026**	0.003
Orthographic uniqueness point	− 0.241***	0.221***	− 0.067***	0.090***
Phonological uniqueness point	0.015	− 0.036	− 0.043***	0.022†
Orthographic spread	− 0.046*	0.031	− 0.011	0.030**
Phonological spread	0.010	− 0.030	0.046***	− 0.016
OLD20	0.058	− 0.023	0.052***	− 0.145***
PLD20	− 0.045	0.024	0.102***	0.001
OLDF	0.167***	− 0.173***	0.091***	− 0.095***
PLDF	0.072**	− 0.078**	0.103***	− 0.100***
Adjusted R^2	**0.403**	**0.273**	**0.558**	**0.346**
ΔR^2	**0.400*****	**0.271*****	**0.504*****	**0.345*****
Spellings and pronunciations				
Spelling (US/UK)	0.069***	− 0.052***	0.016***	− 0.008
Homophone	0.012	− 0.003	− 0.004	0.004
Adjusted R^2	**0.408**	**0.275**	**0.558**	**0.346**
ΔR^2	**0.005*****	**0.003****	**0.000*****	**0.000**
Syllable 1 consistency variables				
Feedforward onset consistency	− 0.011	0.001	0.014*	− 0.006
Feedforward rime consistency	− 0.068***	0.055***	0.008	0.008
Feedback onset consistency	0.026†	− 0.022	− 0.033***	0.010
Feedback rime consistency	− 0.005	0.017	0.026***	− 0.018*
Adjusted R^2	**0.412**	**0.278**	**0.559**	**0.348**
ΔR^2	**0.005*****	**0.004****	**0.001*****	**0.001*****
Composite consistency variables				
Feedforward onset consistency			− 0.051***	0.046***
Feedforward rime consistency			− 0.066***	0.056***
Feedback onset consistency			0.041***	− 0.019**
Feedback rime consistency			− 0.020**	0.012
Adjusted R^2			**0.563**	**0.350**
ΔR^2	NA	NA	**0.004*****	**0.003*****

*p < 0.05, **p < 0.01, ***p < 0.001, †p < 0.10

Table 4. Standardized RT and accuracy regression coefficients from Steps 1 to 6 of the item-level regression analyses for visual lexical decision performance for monosyllabic words and multisyllabic words from the ELP. The p value for each coefficient and R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words ($n = 5714$)		Multisyllabic words ($n = 31{,}268$)	
	RT	Accuracy	RT	Accuracy
Surface variables (onsets)				
Adjusted R^2	**0.005*****	**0.006*****	**0.013*****	**0.003*****
Surface variables (stress)				
Adjusted R^2			**0.084**	**0.003**
ΔR^2	**NA**	**NA**	**0.071*****	**0.001*****
Standard lexical variables				
Number of syllables	NA	NA	0.177***	− 0.133***
Word length	0.209***	− 0.118***	0.205***	0.184***
Number of morphemes	− 0.112***	0.174***	− 0.047***	0.143***
Log_Freq_HAL	− 0.668***	0.613***	− 0.541***	0.564***
Orthographic N	0.025	0.014	0.001	− 0.005
Phonological N	− 0.011	0.001	0.049***	− 0.021*
Orthographic uniqueness point	− 0.236***	0.245***	− 0.083***	0.120***
Phonological uniqueness point	− 0.016	0.019	− 0.042***	0.023*
Orthographic spread	− 0.051***	0.043**	− 0.011†	0.030**
Phonological spread	− 0.033**	− 0.007	0.009	− 0.005
OLD20	0.121***	− 0.097**	0.143***	− 0.188***
PLD20	− 0.057†	0.031	0.114***	− 0.003
OLDF	0.081***	− 0.159***	0.066***	− 0.071***
PLDF	0.045**	− 0.084***	0.064***	− 0.118***
Adjusted R^2	**0.478**	**0.397**	**0.602**	**0.362**
ΔR^2	**0.473*****	**0.392*****	**0.519*****	**0.358*****
Spellings and pronunciations				
Spelling (US/UK)	0.032**	− 0.036**	− 0.005	0.007
Homophone	− 0.018†	0.000	− 0.004	0.007
Adjusted R^2	**0.480**	**0.398**	**0.602**	**0.362**
ΔR^2	**0.001*****	**0.001****	**0.000**	**0.000**
Syllable 1 consistency variables				
Feedforward onset consistency	− 0.039***	0.030**	0.007	0.008
Feedforward rime consistency	− 0.040***	0.042***	0.024***	0.018**
Feedback onset consistency	− 0.013	− 0.010	− 0.040***	0.011†
Feedback rime consistency	− 0.024*	0.005	0.018***	− 0.029***
Adjusted R^2	**0.483**	**0.400**	**0.603**	**0.365**
ΔR^2	**0.004*****	**0.002*****	**0.001*****	**0.003*****
Composite consistency variables				
Feedforward onset consistency			− 0.049***	0.056***
Feedforward rime consistency			− 0.067***	0.058***
Feedback onset consistency			0.044***	− 0.016**
Feedback rime consistency			− 0.033***	0.029***
Adjusted R^2			**0.608**	**0.369**
ΔR^2	**NA**	**NA**	**0.005*****	**0.004*****

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, $^\dagger p < 0.10$

Table 5. Standardized RT and accuracy regression coefficients from Steps 1 to 7 of the item-level regression analyses for auditory lexical decision performance for monosyllabic words and multisyllabic words from the AELP. The p value for each coefficient and R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words ($n = 2256$)		Multisyllabic words ($n = 7531$)	
	RT	Accuracy	RT	Accuracy
Sound file duration				
Adjusted R^2	**0.305*** **	**0.002***	**0.478*** **	**0.001***
Surface variables (onsets)				
Adjusted R^2	**0.353**	**0.009**	**0.494**	**0.004**
ΔR^2	**0.052*** **	**0.012** **	**0.017*** **	**0.005** **
Surface variables (stress)				
Adjusted R^2			**0.494**	**0.007**
ΔR^2	NA	NA	**0.000**	**0.004*** **
Standard lexical variables				
Number of syllables	NA	NA	$-0.032^\dagger$	0.027
Number of phonemes	$0.089^\dagger$	$-0.100^\dagger$	0.255***	-0.039
Number of morphemes	0.063**	-0.140***	0.035***	-0.012
Log_Freq_HAL	-0.249***	0.318***	-0.332***	0.461***
Orthographic N	0.030	0.009	-0.013	0.015
Phonological N	0.078	$0.104^\dagger$	0.070***	-0.020
Orthographic uniqueness point	-0.084**	0.072*	-0.099***	0.076**
Phonological uniqueness point	-0.046	0.169**	-0.031*	0.081***
Orthographic spread	0.001	0.040	0.040**	-0.003
Phonological spread	0.009	0.000	0.004	0.006
OLD20	0.070	-0.025	0.050**	-0.073**
PLD20	-0.024	0.017	0.014	-0.004
OLDF	0.057*	$-0.056^\dagger$	0.087***	-0.044**
PLDF	0.075**	-0.029	0.022*	$-0.029^\dagger$
Adjusted R^2	**0.429**	**0.166**	**0.628**	**0.226**
ΔR^2	**0.079*** **	**0.161*** **	**0.134*** **	**0.220*** **
Spellings and pronunciations				
Spelling (US/UK)	-0.007	$0.037^\dagger$	-0.042***	0.036**
Homophone	0.023	-0.016	0.005	0.000
Adjusted R^2	**0.429**	**0.167**	**0.629**	**0.227**
ΔR^2	**0.001**	**0.001**	**0.002*** **	**0.001** **
Syllable 1 consistency variables				
Feedforward onset consistency	-0.056**	0.067**	0.014	-0.018
Feedforward rime consistency	-0.060***	0.076***	$0.019^\dagger$	0.019
Feedback onset consistency	-0.010	0.032	-0.013	0.027*
Feedback rime consistency	-0.051**	0.075**	0.010	0.012
Adjusted R^2	**0.439**	**0.184**	**0.630**	**0.229**
ΔR^2	**0.010*** **	**0.018*** **	**0.000***	**0.002** **
Composite consistency variables				
Feedforward onset consistency			0.014	0.013
Feedforward rime consistency			-0.004	$0.030^\dagger$
Feedback onset consistency			-0.011	-0.012
Feedback rime consistency			-0.016	0.004
Adjusted R^2			**0.630**	**0.229**
ΔR^2	NA	NA	**0.000**	**0.001**

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, $^\dagger p < 0.10$

Monosyllabic vs. multisyllabic words

For monosyllabic words, syllable 1 consistency measures accounted for a considerable amount of unique variance in response latencies (3.3% in speeded pronunciation, 0.5% (ECP) and 0.4% (ELP) in visual lexical decision, 1.0% in auditory lexical decision), and in response accuracy (2.4% in speeded pronunciation, 0.4% (ECP) and 0.2% (ELP) in visual lexical decision, 1.8% in auditory lexical decision). In contrast, for multisyllabic words, composite consistency measures accounted for more incremental variance in response latencies (1.5% in speeded pronunciation, 0.4% (ECP) and 0.5% (ELP) in visual lexical decision) than first syllable consistency measures (1.0% in speeded pronunciation, 0.3% (ECP) and 0.4% (ELP) in visual lexical decision), although this was limited to only the naming and visual lexical decision tasks. The pattern of results was the same for response accuracy.

Interestingly, for the auditory lexical decision task, neither first syllable nor composite consistency measures appear to account for much variance in both response latencies and response accuracy for multisyllabic words. This is in marked contrast to monosyllabic words, where syllable 1 consistency measures explain a highly significant and substantial amount of variance in both response time and accuracy rates. Previously discussed perspectives of the effects of consistency on spoken word recognition have been based almost unilaterally on monosyllabic words (e.g., Dich, 2011; Petrova et al., 2011; Pattamadilok et al., 2007; Ziegler & Ferrand, 1998; Ziegler et al., 2004, 2008), and the results here are theoretically important because they suggest that findings in the current literature may be specific to only monosyllabic words. In light of this observation, further research should explore whether the processing of multisyllabic words is distinct from monosyllabic words in the auditory lexical decision task, such that consistency exerts less of an influence on the former.

Nonetheless, at least for the speeded pronunciation and lexical decision tasks, consistency effects evidently operate beyond the initial syllable for multisyllabic words. To reiterate, composite consistency reflects the mean consistency across syllabic positions, and since composite consistency accounts for more unique variance than first syllable consistency measures in multisyllabic words, consideration of consistency in all syllabic positions is evidently critical for multisyllabic words.

Feedforward vs. feedback consistency

Focusing on the influence of syllable 1 consistency on the processing of monosyllabic words, and that of composite consistency on multisyllabic words, the effects of feedforward and feedback onset and rime consistency were, wherever significant, facilitatory in nature. In general, more consistent items yielded faster response latencies and higher response accuracies across all tasks.

One peculiar exception to the facilitatory trend was the finding of inhibitory composite feedback *onset* consistency effects in the recognition of multisyllabic words, for both response latencies and response accuracy in the speeded pronunciation and visual lexical decision tasks. Interestingly, Yap and Balota (2009) obtained the exact same findings, and likewise found this observation perplexing. Additionally, composite feedforward *onset* consistency also had inhibitory effects on the response latencies of multisyllabic words in the auditory lexical decision task. In the context of other facilitatory observations, it is difficult to interpret what these results signify, and perhaps future analyses can shed light on plausible explanations for these observations.

Despite the above anomalies, findings for *rime* consistency were relatively stable. For monosyllabic words, first syllable feedforward rime consistency facilitated both response latencies and accuracy in all three tasks, while first syllable feedback rime consistency facilitated both response latencies and accuracy only in the auditory lexical decision task, as well as response latencies in the speeded pronunciation task. For multisyllabic words, both composite feedforward and composite feedback rime consistency facilitated response latencies and accuracy in the speeded pronunciation and visual lexical decision tasks.

Given the conflicting literature on feedback consistency effects in visual word recognition, these findings are especially notable. After partialing out additional predictor variables that may have been confounds in prior studies, these results suggest that feedback consistency effects are not artifactual, and indeed play a role in the visual word recognition of multisyllabic words. Furthermore, there is also striking evidence that feedforward consistency effects are important in predicting auditory lexical decision performance for monosyllabic words. Taken together, these findings provide strong support for the recurrent network theory of word perception (Stone & Van Orden, 1994; Stone et al., 1997; Van Orden & Goldinger, 1994; also see Fig. 1), wherein the bidirectional flow of activation between orthography and phonology facilitates lexical processing, regardless of the modality of stimulus input.

Consistency × frequency interaction

In addition to the main effects of consistency discussed above, the consistency × frequency interaction was also tested. Since the seminal work of Andrews (1982), studies have demonstrated that the effects of atypical spelling-sound correspondences appear to vary with the frequency of the word (e.g., Cortese & Simpson, 2000; Hino & Lupker, 2000; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). Across a variety of tasks in both visual (e.g., Jared, 2002; Lacruz & Folk, 2004; Yap &

Balota, 2009) and spoken (e.g., Petrova et al., 2011) word recognition, researchers have observed that low-frequency words show larger consistency effects than high-frequency words. This interaction has been explained within dual-route models of reading (e.g., Coltheart et al., 2001), as well as connectionist frameworks (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989).

To test the consistency × frequency interaction in this study, first syllable and composite feedforward and feedback consistency × frequency interaction terms were introduced in a final step of the regression analyses. Only rime consistency was included in the interaction terms, as rime consistency effects were the most stable in the foregoing analyses. A summary of the interaction effects is presented in Tables 6, 7, 8, and 9 for speeded pronunciation, visual lexical decision, and auditory lexical decision respectively.

Monosyllabic words For monosyllabic words, first syllable consistency × frequency interaction effects were ambiguous. First syllable feedforward rime consistency × frequency interactions were not significant auditory lexical decision response latencies and accuracy, all ps > 0.05. This interaction was significant in speeded pronunciation response latencies (β = .01, p < 0.001) and response accuracy (β = .00, p = .01), and visual lexical decision response latencies, (ECP: β = -.02, p < .001; ELP: β = -.03, p < 0.001) and response accuracy (ECP: β = 0.00, p = 0.01; ELP: β = 0.01, p = < 0.001), but follow-up analyses revealed that the interactions were not in the expected direction; there were larger consistency effects (i.e., slower response latencies and lower response accuracy) for high-frequency words than low-frequency words.

First syllable feedback rime consistency × frequency interactions had more consistent results; this interaction was significant in response latencies for both speeded pronunciation (β = 0.02, p = < 0.001) and visual lexical decision (ECP: β = 0.01, p = < 0.001; ELP: β = 0.01, p = < 0.001), as well as response accuracy for both speeded pronunciation (β = − 0.01, p = < 0.001) and visual lexical decision (ECP: β = − 0.01, p = < 0.001; ELP: β = − 0.01, p = < 0.001). Specifically, first syllable feedback rime consistency × frequency interactions all showed the expected larger consistency effects for low-frequency words than high-frequency words, although this was limited to only the naming and visual lexical decision tasks. First syllable feedback rime consistency × frequency interactions were not significant in auditory lexical decision response latencies and accuracy, both ps > 0.05.

Multisyllabic words For multisyllabic words, the composite feedforward rime consistency × frequency interaction was significant in response latencies for speeded pronunciation (β = 0 .07, p < 0.001), and visual lexical decision (ECP: β = 0.04, p < 0.001; ELP: β = 0.02, p < 0.001), as well as response accuracy for speeded pronunciation (β = − 0.03, p < 0.001) and visual lexical decision (ECP: β = − 0.01, p < 0.001; ELP: β = − 0.03, p < 0.001). The composite feedback rime consistency × frequency interaction was also significant in response latencies for speeded pronunciation (β = 0.03, p < 0.001) and visual lexical decision (ELP: β = 0.01, p < 0.001), as well as response accuracy for speeded pronunciation (β = − 0.01, p < 0.001) and visual lexical decision (ECP: β = − 0.01, p < 0.001; ELP: β = − 0.02, p <

Table 6. Standardized RT and accuracy regression coefficients for first syllable and composite consistency × frequency interactions (Step 7) in speeded pronunciation performance for monosyllabic words and multisyllabic words from the ELP. The p value for each R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words (n = 5714)		Multisyllabic words (n = 31,268)	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets	R^2 = 0.282***	R^2 = 0.000	R^2 = 0.036***	R^2 = 0.003***
Step 2: Stress typicality	NA	NA	R^2 = 0.125***	R^2 = 0.018***
Step 3: Standard lexical variables	R^2 = 0.531***	R^2 = 0.198***	R^2 = 0.534***	R^2 = 0.246***
Step 4: Spellings and pronunciations	R^2 = 0.533***	R^2 = 0.199**	R^2 = 0.535**	R^2 = 0.246
Step 5: Syllable 1 consistency variables	R^2 = 0.566***	R^2 = 0.223***	R^2 = 0.544***	R^2 = 0.250***
Step 6: Composite consistency variables	NA	NA	R^2 = 0.558***	R^2 = 0.260***
Step 7: Consistency × frequency	R^2 = 0.568***	R^2 = 0.228***	R^2 = 0.565***	R^2 = 0.275***
	beta	beta	beta	beta
Syllable 1 Feedforward rime × frequency	− 0.01***	0.00*	0.00	0.00
Syllable 1 Feedback rime × frequency	0.02***	− 0.01***	− 0.01***	0.01***
Composite Feedforward rime × frequency	NA	NA	0.07***	− 0.03***
Composite Feedback rime × frequency	NA	NA	0.03***	− 0.01***

*p < 0.05, **p < 0.01, ***p < 0.001, †p < 0.10

Table 7. Standardized RT and accuracy regression coefficients for first syllable and composite consistency × frequency interactions (Step 7) in visual lexical decision performance for monosyllabic words and multisyllabic words from the ECP. The *p* value for each R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words (n = 3802)		Multisyllabic words (n = 22,462)	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets	$R^2 = 0.004$**	$R^2 = 0.004$***	$R^2 = 0.008$***	$R^2 = 0.001$***
Step 2: Stress typicality	NA	NA	$R^2 = 0.054$***	$R^2 = 0.001$***
Step 3: Standard lexical variables	$R^2 = 0.403$***	$R^2 = 0.273$***	$R^2 = 0.558$***	$R^2 = 0.346$***
Step 4: Spellings and pronunciations	$R^2 = 0.408$***	$R^2 = 0.275$***	$R^2 = 0.558$***	$R^2 = 0.346$
Step 5: Syllable 1 consistency variables	$R^2 = 0.412$***	$R^2 = 0.278$***	$R^2 = 0.559$***	$R^2 = 0.348$***
Step 6: Composite consistency variables	NA	NA	$R^2 = 0.563$***	$R^2 = 0.350$***
Step 7: Consistency × frequency	$R^2 = 0.415$***	$R^2 = 0.281$***	$R^2 = 0.566$***	$R^2 = 0.354$***
	beta	*beta*	*beta*	*beta*
Syllable 1 Feedforward rime × frequency	− 0.02***	0.00*	0.00†	0.00
Syllable 1 Feedback rime × frequency	0.01***	− 0.01***	− 0.01***	0.00***
Composite Feedforward rime × frequency	NA	NA	0.04***	− 0.01***
Composite Feedback rime × frequency	NA	NA	0.00	− 0.01***

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, †$p < 0.10$

0.001). As expected, these composite consistency × frequency interactions all showed larger consistency effects for low-frequency words than high-frequency words. However, this pattern of results was once again not replicated in the auditory lexical decision task; none of the consistency × frequency interaction effects were significant in

auditory lexical decision response latencies and accuracy, all *p*s > 0.05.

Collectively, these results suggest that the consistency × frequency interaction is much more robust in multisyllabic words than monosyllabic words, for both the speeded pronunciation and visual lexical decision tasks. Given

Table 8. Standardized RT and accuracy regression coefficients for first syllable and composite consistency × frequency interactions (Step 7) in visual lexical decision performance for monosyllabic words and multisyllabic words from the ELP. The *p* value for each R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words (n = 5714)		Multisyllabic words (n = 31,268)	
	RT	Accuracy	RT	Accuracy
Step 1: Phonological onsets	$R^2 = 0.005$***	$R^2 = 0.006$***	$R^2 = 0.013$***	$R^2 = 0.003$***
Step 2: Stress typicality	NA	NA	$R^2 = 0.084$***	$R^2 = 0.003$***
Step 3: Standard lexical variables	$R^2 = 0.478$***	$R^2 = 0.397$***	$R^2 = 0.602$***	$R^2 = 0.362$***
Step 4: Spellings and pronunciations	$R^2 = 0.480$***	$R^2 = 0.398$**	$R^2 = 0.602$	$R^2 = 0.362$
Step 5: Syllable 1 consistency variables	$R^2 = 0.483$***	$R^2 = 0.400$***	$R^2 = 0.603$***	$R^2 = 0.365$***
Step 6: Composite consistency variables	NA	NA	$R^2 = 0.608$***	$R^2 = 0.369$***
Step 7: Consistency × frequency	$R^2 = 0.487$***	$R^2 = 0.402$***	$R^2 = 0.609$***	$R^2 = 0.377$***
	beta	*beta*	*beta*	*beta*
Syllable 1 Feedforward rime × frequency	− 0.03***	0.01***	0.00	0.00***
Syllable 1 Feedback rime × frequency	0.01***	− 0.01***	− 0.01*	0.01*
Composite Feedforward rime × frequency	NA	NA	0.02***	− 0.03***
Composite Feedback rime × frequency	NA	NA	0.01***	− 0.02***

*$p < 0.05$, **p < 0.01, ***p < 0.001, †$p < 0.10$

Table 9. Standardized RT and accuracy regression coefficients for first syllable and composite consistency × frequency interactions (Step 8) in auditory lexical decision performance for monosyllabic words and multisyllabic words from the AELP. The p value for each R^2 change is represented with *asterisks*

Predictor variable	Monosyllabic words (n = 2256)		Multisyllabic words (n = 7531)	
	RT	Accuracy	RT	Accuracy
Step 1: Sound file duration	R^2 = 0.305	R^2 = 0.002	R^2 = 0.478***	R^2 = 0.001*
Step 2: Phonological onsets	R^2 = 0.353	R^2 = 0.009	R^2 = 0.494***	R^2 = 0.004**
Step 3: Stress typicality	NA	NA	R^2 = 0.494	R^2 = 0.007***
Step 4: Standard lexical variables	R^2 = 0.429	R^2 = 0.166	R^2 = 0.628***	R^2 = 0.226***
Step 5: Spellings and pronunciations	R^2 = 0.429	R^2 = 0.167	R^2 = 0.629***	R^2 = 0.227**
Step 6: Syllable 1 consistency variables	R^2 = 0.439	R^2 = 0.184	R^2 = 0.630*	R^2 = 0.229**
Step 7: Composite consistency variables	NA	NA	R^2 = 0.630	R^2 = 0.229
Step 8: Consistency × frequency	R^2 = 0.439	R^2 = 0.184	R^2 = 0.630	R^2 = 0.229
	beta	beta	beta	beta
Syllable 1 Feedforward rime × frequency	0.01	-0.01	0.00	0.00
Syllable 1 Feedback rime × frequency	– 0.02[†]	0.00	0.00	0.00
Composite Feedforward rime × frequency	NA	NA	– 0.01	0.01
Composite Feedback rime × frequency	NA	NA	0.00	0.00

*p < 0.05, **p < 0.01, ***p < 0.001, †p < 0.10

that extant theories on this interaction in visual word recognition have largely been informed by studies on monosyllabic words, the inconsistent findings here are difficult to explain. However, it should be noted that several factorial studies that used only monosyllabic stimuli also failed to observe a significant rime consistency × frequency interaction *by items* for both naming and visual lexical decision reaction times and error rates (Andrews, 1982; Jared, 1997, 2002; Lacruz & Folk, 2004). Likewise, Treiman et al. (1995) did not find a significant rime consistency × frequency interaction in their regression analyses for two separate large sets of monosyllabic visual lexical decision and naming response latencies, after all standard lexical covariates and consistency variables had been controlled for. Hence, although the results for monosyllabic words here are puzzling, they suggest that the consistency × frequency interaction effect may be less reliable in monosyllabic words than multisyllabic words.

More intriguingly, none of the consistency × frequency interactions were significant in the auditory lexical decision task, for both monosyllabic words and multisyllabic words response latencies and accuracy. The consistency × frequency interaction effect has rarely been examined in spoken word recognition (Petrova et al., 2011), and these results are concordant with Pattamadilok et al. (2007) in suggesting that the effect could be less readily observed in auditory lexical decision than in the visual lexical processing tasks. Future work can be directed towards examining whether the interaction could have been modulated

by the specific demands of an auditory lexical decision task, resulting in null effects.

Consistency for multisyllabic words

The consistency norms computed for multisyllabic words here were derived by averaging across syllables. This method operationalizes consistency within individual syllables, and assumes that syllables mediate lexical access. However, as discussed in the introduction, there have been a few other approaches to defining consistency for multisyllabic words. Most prominently, Chateau and Jared (2003) computed body-of-the-BOSS (BOB) consistency norms for disyllabic words, in line with Taft's (1992) theory that the main body of a multisyllabic word is a higher-order orthographic segment that comprises first syllable nucleus, and all consonants following the vowel that would make a orthographically legal combination (e.g., "-ind" in "window", and "-und" in "thunder"; see also Taft, 1979, for Basic Orthographic Syllable Structure (BOSS)). Chateau and Jared (2003) found that BOB consistency was a significant predictor of naming latencies and accuracies, suggesting that consistency can also operate beyond syllabic boundaries.

In addition, one might also argue that it is possible for performance on lexical processing tasks to be predicted by the least consistent or most consistent syllable in a multisyllabic word, instead of the average consistency across syllables. As such, in order to test the predictive power of these

alternative measures, additional regression models were run for multisyllabic words. The hierarchy of the regression models was essentially the same as the preceding main analyses, with the exception that the final step of each model was replaced with either the lowest feedforward and feedback onset and rime values (lowest consistency), highest feedforward and feedback onset and rime values (highest consistency), or feedforward and feedback first syllable onset and rime consistency, but with rime consistency conditioned upon second syllable onset[4]. First syllable consistency measures were omitted from the models due to high correlations with the conditional probability measure.

Table 10 compares the amount of variance predicted by these measures against the composite consistency measures. It is evident that the composite consistency measures clearly outperform other possible measures of consistency for multisyllabic words, in terms of amount of variance predicted in the speeded pronunciation and visual lexical decision tasks. For the auditory lexical decision task, however, it appears that other measures may explain more variance in response latencies and accuracies than the composite consistency measure, although this difference is modest (0.1%).

It is worth noting that the results of this comparison do not necessarily imply that words are always parsed into syllables during lexical processing, but rather, considering the length of multisyllabic words, composite consistency is a better reflection of what constitutes consistency in longer words. Given the limited effects of consistency on auditory lexical decision found in the main analyses, future research can also consider additional approaches to exploring consistency effects in spoken word recognition.

Conclusions

To summarize, the norms presented herein comprise consistency measures for 37,677 words in the ELP, and constitute the largest database of English consistency norms to date. This database goes beyond existing datasets (e.g., Yap & Balota, 2009; Ziegler et al., 1997b) by providing feedforward and feedback consistency data for both sub-syllabic and composite measures, thus encompassing a full range of consistency operationalized at various levels. Such a resource has been lacking for previous studies that needed to deal with the diverse range of variables influencing word recognition, and should hopefully serve as a resource to researchers in their selection of stimuli.

Using both correlation and regression analyses, the effects of consistency on speeded pronunciation and both visual and auditory lexical decision performance were also empirically examined. Consistency generally facilitated response latencies and accuracy, thus replicating previous research and establishing the validity of the present norms. Several observations regarding the effects of consistency on lexical processing were also made: (a) consistency accounts for more incremental variance in speeded pronunciation, compared to both lexical decision tasks; (b) composite consistency is a more appropriate index of consistency in multisyllabic words, rather than first syllable consistency measures; (c) both feedforward and feedback consistency effects are stable predictors of both visual and spoken word recognition performance; (d) the consistency × frequency interaction effect is much more robust in multisyllabic words than monosyllabic words for the speeded pronunciation and visual lexical decision tasks; and (e) consistency effects in multisyllabic words and the consistency × frequency interactions are not reliable in auditory lexical decision, highlighting the need for further research on the effects of consistency on spoken word recognition.

It is worth noting that analyses were only completed for first syllable and composite onset and rime consistency measures. In this respect, it remains to be seen how other sub-syllabic measures for later syllables would perform in similar regression models. Also, the analyses here focused on the influence of consistency on visual and spoken word recognition, but from a broader theoretical perspective, these norms will also be useful for strengthening and developing new lines of research in measures of lexical processing beyond the prototypical lexical decision and naming tasks, such as spelling performance.

Final remarks

As is the case with most language-specific normative material, it is important to recognize that the consistency norms were computed based on American English spelling and pronunciation, and caution should be exercised if there are intentions to utilize this set of measures for stimuli in other dialectical variations (e.g., British English). In the same vein, the fact that the computations were based on an adult corpus might restrict the generalizability of the measures to studies on different demographics, such as developing readers.

Nonetheless, this is the first exhaustive database of consistency norms to be developed, and as such, presents a major step forward in the field's understanding of the influence of spelling-sound consistency on lexical processing. From a methodological perspective, these norms will be a useful tool for researchers looking to effectively control and manipulate lexical variables in their studies,

---

[4] Although this conditional probability measure does not exactly capture the BOSS structure and BOB consistency, it is based on the same underlying assumption that extrasyllabic contextual information from second syllable onset can help constrain the pronunciation of the rime in the first syllable.

Table 10. Amount of variance explained (R^2) by composite consistency, lowest consistency, highest consistency, and first syllable onset with rime consistency conditioned upon second syllable onset measures (Steps 5–6) in speeded pronunciation, and visual and auditory lexical decision performance for multisyllabic words. The p value for each R^2 change is represented with *asterisks*. The largest ΔR^2 is in bold

Predictor variable	Composite consistency		Lowest consistency		Highest consistency		S1 rime | S2 onset consistency	
	RT	Accuracy	RT	Accuracy	RT	Accuracy	RT	Accuracy
Speeded Pronunciation (ELP), n = 31,268								
Step 1: Phonological onsets	$R^2 = 0.036$***	$R^2 = 0.003$***	$R^2 = 0.036$***	$R^2 = 0.003$***	$R^2 = 0.036$***	$R^2 = 0.003$***	$R^2 = 0.036$***	$R^2 = 0.003$***
Step 2: Stress typicality	$R^2 = 0.125$***	$R^2 = 0.018$***	$R^2 = 0.125$***	$R^2 = 0.018$***	$R^2 = 0.125$***	$R^2 = 0.018$***	$R^2 = 0.125$***	$R^2 = 0.018$***
Step 3: Standard lexical variables	$R^2 = 0.534$***	$R^2 = 0.246$***	$R^2 = 0.534$***	$R^2 = 0.246$***	$R^2 = 0.534$***	$R^2 = 0.246$***	$R^2 = 0.534$***	$R^2 = 0.246$***
Step 4: Spellings and pronunciations	$R^2 = 0.535$**	$R^2 = 0.246$	$R^2 = 0.535$**	$R^2 = 0.246$	$R^2 = 0.535$**	$R^2 = 0.246$	$R^2 = 0.535$**	$R^2 = 0.246$
Step 5: Consistency	**$R^2 = 0.556$*****	**$R^2 = 0.259$*****	$R^2 = 0.550$***	$R^2 = 0.254$***	$R^2 = 0.540$***	$R^2 = 0.250$***	$R^2 = 0.546$***	$R^2 = 0.253$***
Visual Lexical Decision (ECP), n = 22,462								
Step 1: Phonological onsets	$R^2 = 0.008$***	$R^2 = 0.001$***	$R^2 = 0.008$***	$R^2 = 0.001$***	$R^2 = 0.008$***	$R^2 = 0.001$***	$R^2 = 0.008$***	$R^2 = 0.001$***
Step 2: Stress typicality	$R^2 = 0.054$***	$R^2 = 0.001$***	$R^2 = 0.054$***	$R^2 = 0.001$***	$R^2 = 0.054$***	$R^2 = 0.001$***	$R^2 = 0.054$***	$R^2 = 0.001$***
Step 3: Standard lexical variables	$R^2 = 0.558$***	$R^2 = 0.346$***	$R^2 = 0.558$***	$R^2 = 0.346$***	$R^2 = 0.558$***	$R^2 = 0.346$***	$R^2 = 0.558$***	$R^2 = 0.346$***
Step 4: Spellings and pronunciations	$R^2 = 0.558$***	$R^2 = 0.346$	$R^2 = 0.558$***	$R^2 = 0.346$	$R^2 = 0.558$***	$R^2 = 0.346$	$R^2 = 0.558$***	$R^2 = 0.346$
Step 5: Consistency	**$R^2 = 0.562$*****	**$R^2 = 0.350$*****	$R^2 = 0.560$***	$R^2 = 0.348$***	$R^2 = 0.561$***	$R^2 = 0.348$***	$R^2 = 0.560$***	$R^2 = 0.348$***
Visual Lexical Decision (ELP), n = 31,268								
Step 1: Phonological onsets	$R^2 = 0.013$***	$R^2 = 0.003$***	$R^2 = 0.013$***	$R^2 = 0.003$***	$R^2 = 0.013$***	$R^2 = 0.003$***	$R^2 = 0.013$***	$R^2 = 0.003$***
Step 2: Stress typicality	$R^2 = 0.084$***	$R^2 = 0.003$***	$R^2 = 0.084$***	$R^2 = 0.003$***	$R^2 = 0.084$***	$R^2 = 0.003$***	$R^2 = 0.084$***	$R^2 = 0.003$***
Step 3: Standard lexical variables	$R^2 = 0.602$***	$R^2 = 0.362$***	$R^2 = 0.602$***	$R^2 = 0.362$***	$R^2 = 0.602$***	$R^2 = 0.362$***	$R^2 = 0.602$***	$R^2 = 0.362$***
Step 4: Spellings and pronunciations	$R^2 = 0.602$	$R^2 = 0.362$	$R^2 = 0.602$	$R^2 = 0.362$	$R^2 = 0.602$	$R^2 = 0.362$	$R^2 = 0.602$	$R^2 = 0.362$
Step 5: Consistency	**$R^2 = 0.607$*****	**$R^2 = 0.368$*****	$R^2 = 0.605$***	$R^2 = 0.366$***	$R^2 = 0.605$***	$R^2 = 0.364$***	$R^2 = 0.604$***	$R^2 = 0.365$***
Auditory Lexical Decision (AELP), n = 7531								
Step 1: Sound file duration	$R^2 = 0.478$***	$R^2 = 0.001$*	$R^2 = 0.478$***	$R^2 = 0.001$*	$R^2 = 0.478$***	$R^2 = 0.001$*	$R^2 = 0.478$***	$R^2 = 0.001$*
Step 2: Phonological onsets	$R^2 = 0.494$***	$R^2 = 0.004$**	$R^2 = 0.494$***	$R^2 = 0.004$**	$R^2 = 0.494$***	$R^2 = 0.004$**	$R^2 = 0.494$***	$R^2 = 0.004$**
Step 3: Stress typicality	$R^2 = 0.494$	$R^2 = 0.007$***	$R^2 = 0.494$	$R^2 = 0.007$***	$R^2 = 0.494$	$R^2 = 0.007$***	$R^2 = 0.494$	$R^2 = 0.007$***
Step 4: Standard lexical variables	$R^2 = 0.628$***	$R^2 = 0.226$***	$R^2 = 0.628$***	$R^2 = 0.226$***	$R^2 = 0.628$***	$R^2 = 0.226$***	$R^2 = 0.628$***	$R^2 = 0.226$***
Step 5: Spellings and pronunciations	$R^2 = 0.629$***	$R^2 = 0.227$**	$R^2 = 0.629$***	$R^2 = 0.227$**	$R^2 = 0.629$***	$R^2 = 0.227$**	$R^2 = 0.629$***	$R^2 = 0.227$**
Step 6: Consistency	$R^2 = 0.629$	$R^2 = 0.228$**	**$R^2 = 0.630$****	$R^2 = 0.228$	**$R^2 = 0.630$***	**$R^2 = 0.229$****	**$R^2 = 0.630$***	$R^2 = 0.228$*

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. †$p < 0.10$

and future empirical work using these norms can help further enrich the understanding of processes involved in lexical processing.

Open practices statement The full set norms are available as supplementary material. None of the data reported here were preregistered in an independent, institutional registry.

Supplementary Material

The full set of norms are provided in an Excel Binary Workbook appended to this paper. For each of the 37,677 words, the file contains the following information: the orthography (Word), number of letters (Length), number of phonemes (N_Phon), number of syllables (N_Syll), the type and token feedforward and feedback onset, nucleus, coda, oncleus, and rime consistency for individual syllables, and the composite measures for multisyllabic words. The phonology of the words were derived from Oxford Dictionaries through application programming interface (API), via a researcher license key granted to the first author. The pronunciations are therefore proprietary to Oxford Dictionaries and cannot be provided.

Appendix

Table 11. Phoneme-grapheme correspondences used for orthographic parsing

Phoneme	Possible graphemes
/b/	b
/d/	d, t
/dʒ/	j, g, d
/ð/	th
/f/	f, ph, gh
/g/	g, x
/h/	h, wh
/j/	y, u
/k/	k, c, q, x
/l/	l, le
/m/	m
/n/	n, kn, gn
/ŋ/	n, ng
/p/	p
/r/	r, wr

Table 11. (continued)

Phoneme	Possible graphemes
/s/	s, c, ps, ts
/ʃ/	sh, t, ch, c, s, sc, ce, su
/t/	t, pt
/tʃ/	ch, c, t
/θ/	th
/v/	v
/w/	w
/z/	z, s, x
/ʒ/	s, su, t, g, z

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*(9), 814–823.

Alegria, J., & Mousty, P. (1996). The development of spelling procedures in French-speaking, normal and reading-disabled children: Effects of frequency and lexicality. *Journal of Experimental Child Psychology, 63*(2), 312–338.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent?. *Memory & Cognition, 10*(6), 565–575.

Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *The Quarterly Journal of Experimental Psychology, 59*(5), 920–948.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*(2), 283–316.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459.

Beyermann, S., & Penke, M. (2014). The impact of orthographic consistency on German spoken word identification. *International Journal of Disability, Development and Education, 61*(3), 212–224.

Bloomfield, L. (1933). *Language.* University of Chicago Press: Chicago.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers, 30*(2), 272–277.

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology, 92*(4), 307–321.

Chateau, D., & Jared, D. (2003). Spelling–sound consistency effects in disyllabic word naming. *Journal of Memory and Language, 48*(2), 255–280.

Clements, G. N., & Keyser, S. J. (1983). CV phonology: A generative theory of the syllable. *Linguistic Inquiry Monograph* (pp. 1–191). Cambridge, Massachusetts: MIT Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Cortese, M. J., & Simpson, G. B. (2000). Regularity effects in word naming: What are they?. *Memory & Cognition, 28*(8), 1269–1276.

Dalby, J. M. (1986). *Phonetic structure of fast speech in American English (Vol. 7).* Reproduced by the Indiana University Linguistics Club.

Davies, R. A., & Weekes, B. S. (2005). Effects of feedforward and feedback consistency on reading and spelling in dyslexia. *Dyslexia, 11*(4), 233–252.

De Cara, B., & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers, 34*(3), 416–423.

Dich, N. (2011). Individual differences in the size of orthographic effects in spoken word recognition: The role of listeners' orthographic skills. *Applied Psycholinguistics, 32*(1), 169–186.

Dich, N. (2014). Orthographic consistency affects spoken word recognition at different grain-sizes. *Journal of Psycholinguistic Research, 43*(2), 141–148.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*(4), 674–691.

Goh, W. D., Yap, M. J., & Chee, Q. W. (in press). The Auditory English Lexicon Project: a multi-talker, multi-region psycholinguistic database of 10,170 spoken words and nonwords. *Behavior Research Methods.* doi: 10.3758/s13428-020-01352-0

Halle, M., & Idsardi, W. (1997). /r/, hypercorrection, and the elsewhere condition. *Derivations and Constraints in Phonology,* 331–348.

Hammond, M. (1999). *The Phonology of English: A Prosodic Optimality-Theoretic Approach: A Prosodic Optimality-Theoretic Approach.* Oxford University Press, UK.

Heselwood, B. (2007). Schwa and the phonotactics of RP English. *Transactions of the Philological Society, 105*(2), 148–187.

Hino, Y., & Lupker, S. J. (2000). Effects of word frequency and spelling–to–sound regularity in naming with and without preceding lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 166.

Hooper, J. B. (1978). Constraints on Schwa-deletion in American. *Recent Developments in Historical Phonology, 4,* 183.

Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36*(4), 505–529.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46*(4), 723–750.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and. Language, 29*(6), 687–715.

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 92.

Jensen, J. T. (2000). Against ambisyllabicity. *Phonology, 17*(2), 187–235.

Kahn, D. (2015). *Syllable-based generalizations in English phonology.* Routledge: New York.

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing, 25*(6), 1259–1282.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology-Amsterdam, 94,* 67.

Katz, L., Lee, C. H., Tabor, W., Frost, S. J., Mencl, W. E., Sandak, R., Rueckl, J., & Pugh, K. R. (2005). Behavioral and neurobiological effects of printed word repetition in lexical decision and naming. *Neuropsychologia, 43*(14), 2068–2083.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language, 37*(3), 295–311.

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language, 44*(4), 592–617.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47*(1), 145–171.

Kessler, B., Treiman, R., & Mullennix, J. (2008). Feedback consistency effects in single-word reading. *Single-word reading: Behavioral and biological perspectives,* 159–174.

Lacruz, I., & Folk, J. R. (2004). Feedforward and feedback consistency effects for high- and low-frequency words in lexical decision and naming. *The Quarterly Journal of Experimental Psychology Section A, 57*(7), 1261–1284.

Lété, B., Peereman, R., & Fayol, M. (2008). Consistency and word-frequency effects on spelling among first-to fifth-grade French children: A regression-based study. *Journal of Memory and Language, 58*(4), 952–977.

Mandera, P., Keuleers, E., & Brysbaert, M. (in press). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods.*

Massaro, D. W., & Jesse, A. (2005). The magic of reading: Too many influences for quick and easy explanations. From orthography to pedagogy: Essays in honor of Richard L. Venezky, 37–61.

The OED and Oxford Dictionaries, Oxford English Dictionary (2020). Retrieved from: https://www.oed.com/page/oedodo/The+OED+and+Oxford+Dictionaries

Oxford Dictionaries, Oxford University Press: Oxford, United Kingdom. (2020). Retrieved from: https://www.lexico.com/

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language and Cognitive Processes, 22*(5), 700–726.

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica, 60*(1), 45–69.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and language, 39*(2), 151–174.

Peereman, R., Dufour, S., & Burt, J. S. (2009). Orthographic influences in spoken word recognition: the consistency effect in semantic and gender categorization tasks. *Psychonomic Bulletin & Review, 16*(2), 363–368.

Perre, L., Pattamadilok, C., Montant, M., & Ziegler, J. C. (2009). Orthographic effects in spoken language: on-line activation or phonological restructuring?. *Brain Research, 1275,* 73–80.

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain Research, 1188,* 132–138.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review, 114*(2), 273.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61*(2), 106–151.

Petrova, A., Gaskell, G., & Ferrand, L. (2011). Orthographic consistency and word-frequency effects in auditory word recognition: New evidence from lexical decision and rime detection. *Frontiers in Psychology, 2,* 263.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115.

Python Software Foundation. Python, version 3.6.3. Available at http://www.python.org

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition?. *Journal of Verbal Learning and Verbal Behavior, 23*(3), 383–404.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8*(6), 411–416.

Stone, G. O., & Van Orden, G. C. (1994). Building a resonance framework for word recognition using design and system principles. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1248.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*(3), 337–359.

Taft, M. (1979). Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Memory and Language, 18*(1), 21–39.

Taft, M. (1992). The body of the BOSS: Subsyllabic units in the lexical processing of polysyllabic words. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1004–1014.

Treiman, R., & Kessler, B. (1995). In defense of an onset-rime syllable structure for English. *Language and Speech, 38*(2), 127–142.

Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88*(1), 49–78.

Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology, 93*(1), 1–24.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124*(2), 107–136.

Van Oostendorp, M. (1995). *Vowel quality and syllable projection.* Tilburg: Katholieke Universiteit Brabant.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1269.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin.* Berlin: Mouton de Gruyter.

Ventura, P., Kolinsky, R., Pattamadilok, C., & Morais, J. (2008). The developmental turnpoint of orthographic consistency effects in speech recognition. *Journal of Experimental Child Psychology, 100*(2), 135–145.

Ventura, P., Morais, J., & Kolinsky, R. (2007). The development of the orthographic consistency effect in speech recognition: From sublexical to lexical involvement. *Cognition, 105*(3), 547–576.

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes, 19*(1), 57–95.

Weekes, B. S., Castles, A. E., & Davies, R. A. (2006). Effects of consistency and age of acquisition on reading and spelling among developing readers. *Reading and Writing, 19*(2), 133–169.

Yap, M. J. (2007). *Visual word recognition: Explorations of megastudies, multisyllabic words, and individual differences* (Unpublished doctoral dissertation). Washington University in St. Louis, St. Louis, Missouri.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*(4), 502–529.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science, 21*(4), 551–559.

Ziegler, J. C. & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683–689.

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition, 32*(5), 732–741.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997a). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37*(4), 533–554.

Ziegler, J. C., Petrova, A., & Ferrand, L. (2008). Feedback consistency effects in visual and auditory word recognition: where do we stand after more than a decade?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(3), 643–661.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997b). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers, 29*(4), 600–618.

Zwicky, A. (1972). On casual speech. In *Eighth Regional Meeting of the Chicago Linguistic Society* (Vol. 8, pp. 607–615).