Visual word recognition of multisyllabic words

Melvin J. Yap [a,*], David A. Balota [b]

[a] Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4, #02-07, Singapore 117570, Republic of Singapore
[b] Department of Psychology, Washington University in St. Louis, MO 63130, United States

ARTICLE INFO

ABSTRACT

The visual word recognition literature has been dominated by the study of *monosyllabic* words in factorial experiments, computational models, and megastudies. However, it is not yet clear whether the behavioral effects reported for monosyllabic words generalize reliably to *multisyllabic* words. Hierarchical regression techniques were used to examine the effects of standard variables (phonological onsets, stress pattern, length, orthographic N, phonological N, word frequency) and additional variables (number of syllables, feedforward and feedback phonological consistency, novel orthographic and phonological similarity measures, semantics) on the pronunciation and lexical decision latencies of 6115 monomorphemic multisyllabic words. These predictors accounted for 61.2% and 61.6% of the variance in pronunciation and lexical decision latencies, respectively, higher than the estimates reported by previous monosyllabic studies. The findings we report represent a well-specified set of benchmark phenomena for constraining nascent multisyllabic models of English word recognition.

Introduction

Understanding the processes underlying the visual recognition of isolated words remains a central endeavor in psycholinguistics, cognitive psychology, and cognitive neuroscience. Over the past three decades, a prodigious amount of work in visual word recognition has not only identified the many statistical properties associated with words (e.g., length, frequency of occurrence, concreteness, etc.) but also the effect of these properties on word recognition performance (see Balota, Yap, & Cortese, 2006, for a review). Importantly, the effects uncovered by this kind of empirical work have also been used to constrain computational models of word recognition.

Extant models themselves have been developed from two different standpoints. The traditional perspective holds that word recognition involves rules operating on explicit local representations (e.g., the dual route cascaded model; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), while the connectionist approach holds that lexi-

cal processing arises as a result of competitive and cooperative interactions among distributed representations (e.g., the triangle model; Plaut, McClelland, Seidenberg, & Patterson, 1996). Both classes of models were developed to account for performance on two tasks, speeded pronunciation (participants read aloud visually printed words) and lexical decision (participants discriminate between real words and made-up words, e.g., *flirp*, via button presses). As reflected in the modeling efforts, accounting for the behavioral effects observed with these two tasks has become the *de facto* gold standard in visual word recognition research. However, neither task is process-pure; each reflects task-general word identification processes *and* task-specific processes. This makes it particularly important for word recognition researchers to consider the effects of variables across *both* tasks.

From monosyllables to multisyllables

The available literature in visual word recognition research has been dominated by the study of monosyllabic words in computational models, factorial experiments,

---

* Corresponding author. Fax: +65 6 773 1843.
  E-mail address: melvin@nus.edu.sg (M.J. Yap).

and megastudies. The emphasis on monosyllabic words is unsurprising. They are relatively simple stimuli to work with because experimenters do not have to worry about additional processes such as syllabic segmentation, stress assignment, and vowel reduction. For example, English disyllabic words can either be stressed on the first syllable (trochaic stress, e.g., *cancel*) or the second syllable (iambic stress, e.g., *comply*). There are also differences in vowel reduction, a process whereby certain vowels undergo qualitative changes when they occur in an unstressed position (e.g., become replaced by a schwa, *bypass* vs. *compass*). Furthermore, from a practical standpoint, many of the measures available for monosyllabic words, such as phonological consistency (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), familiarity (Balota, Pilotti, & Cortese, 2001), and imageability (Cortese & Fugett, 2004), simply have not been developed for a comprehensive set of multisyllabic words.

Although the past emphasis on monosyllabic words is understandable, it is possible that the results from monosyllabic studies may not necessarily generalize to more complex multisyllabic words. This is important since monosyllabic words only comprise a minority of the words in a person's lexicon. For example, in the English Lexicon Project (ELP; Balota et al., 2007), an on-line repository of lexical and behavioral measures for 40,481 words (http://elexicon.wustl.edu), only about 15% of the 40,000 words represented are monosyllabic. We shall now turn to a selective overview of the multisyllabic word recognition literature.

Multisyllabic word recognition

Phonological consistency

Jared and Seidenberg (1990) examined pronunciation latencies for di- and trisyllabic words that varied with respect to their phonological consistency. At that time, the major theoretical view (e.g., Patterson & Morton, 1985) was that spelling-sound mappings were best represented via a set of abstract grapheme-phoneme correspondence rules (e.g., k → /k/). Under this perspective, words whose pronunciations are rule-governed are *regular* while words that deviate from the rules are *exceptional*. The adequacy of this dichotomy was first challenged by Glushko's (1979) influential study, where he argued that regularity had to be supplemented by *consistency*. Consistency reflects the mapping between spelling and sound, and a word is considered consistent if its pronunciation matches that of most similarly spelled words. For example, *spook* is inconsistent because the pronunciation of its rime (vowel and following consonants, i.e., *ook*) conflicts with that of similarly spelled words (e.g., *book*, *look*, *hook*). Glushko examined exceptional words (e.g., *have*), regular *consistent* words (e.g., *wade*), and regular *inconsistent* words (e.g., *wave*), and reported that pronunciation latencies were longer for both exception and regular inconsistent words, compared to regular consistent words. Jared and Seidenberg (Experiment 1) developed multisyllabic analogs of Glushko's stimuli to explore consistency effects in multisyllabic word recognition. Specifically, *ravine* is exceptional because the second syllable is pronounced differently in

isolation. *Convene* is regular consistent because the second syllable is pronounced the same way in all words with the same stress pattern. In contrast, *divine* is regular inconsistent because although the second syllable is pronounced the same way in isolation, the second syllable can be pronounced two different ways (i.e., *ravine* vs. *divine*). Compared to regular words, exceptional and regular inconsistent words yielded longer pronunciation latencies, mirroring Glushko's results with monosyllabic words. Moreover, this effect was observed only for low-frequency words (Jared and Seidenberg (1990), Experiment 2).

Jared and Seidenberg (1990, Experiment 3) also reported that although words with more syllables took longer to pronounce, this effect was limited to low-frequency words. Since this finding is consistent with low-frequency words being decomposed into syllables during word recognition, a final experiment was conducted which involved syllable-by-syllable presentation of words. In this condition, participants pronounced exception words more slowly and less accurately, compared to when words were presented as wholes. In contrast, the pronunciation of regular words was not affected by the syllabified stimulus presentation. Based on this final finding, Jared and Seidenberg argued that it is unlikely that pronunciations are generated on a syllable-by-syllable basis. If syllabification is indeed occurring, then a stimulus display that makes syllabic units more salient should *not* impair performance. Instead, it is likely that readers are using information beyond the first syllable to constrain the pronunciation of an exception word. When this information is unavailable, pronunciation latencies are slowed down.

Although consistency was initially treated as a dichotomous variable (Glushko, 1979; Jared & Seidenberg, 1990), subsequent studies (e.g., Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995) have defined consistency in a continuous manner, after Jared, McRae, and Seidenberg (1990) demonstrated that the magnitude of consistency effects is related to the relative number of "friends" (i.e., similarly spelled words pronounced the same way) and "enemies" (i.e., similarly spelled words pronounced differently) in a word's neighborhood. Treiman et al. computed spelling-to-sound consistency measures for various orthographic segments (e.g., C_1, V, C_2, C_1V, VC_2 in monosyllabic words), basing them on the number or frequency of friends relative to the total number or frequency of friends *and* enemies. Using regression analyses, these measures were used to predict speeded pronunciation performance in two large databases of monosyllabic CVC words. The major finding was that the consistency of the VC_2 segment (i.e., the orthographic rime) accounted for variance in pronunciation latencies and errors even after lexical variables and the consistency of individual graphemes were controlled for.

Chateau and Jared (2003) extended Treiman et al.'s methodology to disyllabic words by examining consistency effects in a large set of 1000 monomorphemic six-letter disyllabic words, after computing the consistency of various intrasyllabic orthographic segments (i.e., C_1, C_1V_1, V_1, V_1C_2, C_2, C_3, C_3V_2, V_2, V_2C_4; see Chateau & Jared, 2003, Fig. 1) for each word. In addition, Chateau and Jared calculated the consistency for a unit that transcends syllabic
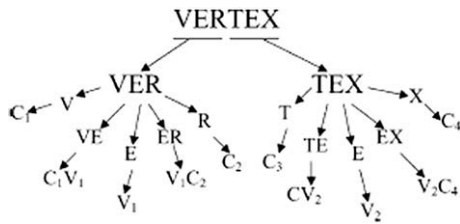
Fig. 1. The division of the word *VERTEX* into different orthographic segments. From Chateau and Jared (2003) p.261. Copyright 2003 by Elsevier. Reproduced with permission.

boundaries, the body of the Basic Orthographic Syllabic Structure (BOSS) (Taft, 1979; Taft, 1992). According to Taft (1979), for multisyllabic words, the BOSS serves as an orthographic access unit to the lexicon. The body of the BOSS, also called the BOB, includes the vowel grapheme of the first syllable and as many consonants that follow the first vowel to form an orthographically legal word ending. For example, for the word *mea-dow*, *ead* is the BOB; this unit straddles the first and second syllables. Taft (1992) has argued that the BOB is to multisyllabic words what the rime is to monosyllabic words, that is, it plays an important role in lexical access.

The analyses conducted by Chateau and Jared (2003) were somewhat complicated by the fact that not all the six-letter disyllabic words selected share the same consonant–vowel structure. That is, only a subset (477) of the 1000 words (e.g., *vertex*) possesses consistency measures for *all* orthographic segments. Other words have structures which lack one or more of these segments. For example, *belong* lacks the C_2 segment while *brandy* lacks the C_4 segment. This made it necessary for Chateau and Jared to run a simultaneous regression with the 477 target words and eight separate hierarchical regression analyses in which different subsets of consistency predictors were entered for words belonging to different structures (since a word is eliminated from a regression analysis if there is a missing value for a predictor). Although this approach is cumbersome, it does have the advantage of assessing the predictive power of a particular orthographic unit in different orthographic structures. Across the various analyses, the spelling-sound consistency of two segments predicted pronunciation performance especially well, V_2 (i.e., second syllable vowel) and BOB consistency. In contrast, the consistency of the consonantal segments did not account for much variance. Interestingly, Chateau and Jared noted that the consistency of the rime segments (i.e., V_1C_2 and the V_2C_4) were relatively weak predictors of pronunciation performance, which seems surprising given previous work showing prominent effects of rime consistency in monosyllabic words (cf. Glushko, 1979, and Treiman et al., 1995). Specifically, V_1C_2 consistency effects were significant in the simultaneous regression analysis ($n = 477$) and three of the eight hierarchical regression analyses, but only in latencies, not in error rates. V_2C_4 consistency effects were similarly limited. A follow-up experiment (Experiment 3) which factorially manipulated consistency and segment (BOB vs. V_1C_2) also yielded significant effects of BOB, but not V_1C_2, consistency. According to Chateau and Jared,

BOB effects may be more robust than V_1C_2 effects because the extra letter in the BOB allows it to better constrain the pronunciation of the vowel in the first syllable. However, it is not the case that pronunciation performance is influenced by *either* V_1C_2 or BOB consistency. It is more likely that the consistency of multiple grain sizes is *simultaneously* influencing speeded pronunciation performance.

Role of the syllable

Based on the findings summarized above, Chateau and Jared (2003) have argued it is unlikely that multisyllabic words are parsed into syllables, with pronunciations assigned syllable-by-syllable. If so, then one would expect stronger rime consistency effects, since the rime maps onto the syllabic body. One would also *not* expect BOB effects, since the BOB crosses syllabic boundaries. In addition, Jared and Seidenberg (1990) showed that number of syllables only influenced the speeded pronunciation of low-frequency words (Experiment 3) and dividing words into syllables did not shorten pronunciation latencies but instead slowed down the processing of exception words (Experiment 4). These data are difficult to reconcile with the notion that *all* words are segmented into syllables. However, although Chateau and Jared rule out the idea that letter strings are explicitly parsed into syllables, their study provides some provocative evidence that syllables can, along with other units, influence speeded pronunciation performance. Interestingly, as we shall now discuss, the role of syllables in visual word recognition is a surprisingly contentious issue.

There has clearly been some strong support for a critical role of the syllable in language processing. For example, the syllable has played a central role in linguistic theories (Blevins, 1995; Hooper, 1972; Selkirk, 1982). In addition, there is considerable evidence for the reality of syllables in the mental representations of words (see Cutler, Mehler, Norris, & Segui, 1986, for a review). Indeed, Spoehr (1981) has claimed a central role for the syllable in the perception of both speech and print. Furthermore, Content, Kearns, and Frauenfelder (2001) highlighted the role of the syllable in leading models of speech production, based on evidence that it is more accessible to metalinguistic manipulations than other phonological units, and that it appears to be a potential processing unit in visual word recognition. For example, using an illusory conjunction paradigm, Prinzmetal, Treiman, and Rho (1986) reported that participants were more likely to produce illusory conjunctions between letters in the *same* syllable of a word than between letters within different syllables, suggesting that syllables are units of analyses in word perception. In a modified lexical decision task, Ashby and Martin (2008) observed that lexical decision latencies were faster when a target word (e.g., *gender*) was primed by a parafoveal preview matching the word's initial syllable (e.g., *gen*) than a preview that contained one letter more or less than the initial syllable (e.g., *ge*).

On the other hand, there are also studies that cast doubt on the view that syllables play an important role. For example, Ferrand, Segui, and Humphreys (1997), using brief masked primes, observed syllable priming effects for English words with clear syllabic boundaries, i.e., faster pronunciation for *bal-cony* when primed by *bal* than when

primed by *ba*. However, it should be noted that priming effects were absent for ambisyllabic words (e.g., *balance*) and in lexical decision performance, suggesting that the role of syllables is restricted to syllabifiable words, and that priming may reflect *output* more than *access* processes. Other linguists have assigned a marginal role to syllables (Chomsky & Halle, 1968) and psychologists have argued that there is little evidence for their functional status, at least in English (see Cutler et al., 1986; Jared & Seidenberg, 1990). Seidenberg (1987), who replicated and extended the Prinzmetal et al. (1986) study, argued that the apparent perceptual salience of syllables in that study was due to orthographic redundancy rather than to syllabification *per se* (note however, that Rapp, 1992, extended Seidenberg's study and found clear effects of syllabic structure that could not be explained away by orthographic redundancy). The robustness of the syllable priming effect itself has also been criticized. For example, Schiller (2000), using the same materials and methodology as Ferrand et al. (1997), was unable to replicate the syllable priming effect in English (Brand, Rey, & Peereman, 2003, were also unable to replicate the effect in French).

The work by Chateau and colleagues helps inform this debate. In Chateau and Jared (2003), although V_2 and BOB consistency effects were more robust than V_1C_2 and V_2C_4 consistency effects, the rime effects were nonetheless reliable, most notably in the simultaneous regression analyses. These results indicate that syllabically-defined (i.e., rime) consistency accounts for unique variance in pronunciation performance, especially when one considers performance on a large set of English words. As we have discussed earlier, participants may be sensitive to both syllables *and* units larger than the syllable (cf. Chateau and Jared). Another finding that bears on the role of the syllable is the number of syllables effect. The underlying assumption here, of course, is that this effect is a marker for processes that recover syllabic structure (Jared & Seidenberg, 1990). As described earlier, Jared and Seidenberg observed that for low-frequency words, words with more syllables took longer to pronounce, after controlling for factors such as length, initial phoneme, syllabic structure, and stress. Likewise, New, Ferrand, Pallier, and Brysbaert (2006), who carried out regression analyses of 33,006 words from the ELP (Balota et al., 2007), reported a positive correlation between number of syllables and lexical decision latencies, after controlling for a number of variables (see Butler & Hains, 1979, who also reported this pattern). Jared and Seidenberg have argued that inferring syllabic decomposition from syllabic structure effects is equivocal because number of syllables, by definition, is confounded with number of vowels, and number of syllables effects may simply reflect additional time devoted to processing vowels, which are the greatest source of inconsistency in most words. Given that words with more syllables tend to be associated with greater feedforward inconsistency, it is critical to control for phonological consistency, which was not done in the New et al. study, in large part because there have not been useful estimates of phonological consistency for multisyllabic words. As described below, we will provide such a measure of consistency in the present research.

To recapitulate, we think that syllables do play a role in visual word recognition, in the sense that multiple codes (i.e., phoneme, syllable, BOB) emerge due to the statistical properties of the orthographic-phonological-semantic mappings which mediate lexical access and output processes. The salience and utility of a code may be stimuli- and task-dependent. Although BOB and V_2 consistency were especially powerful predictors across the various regression analyses in Chateau and Jared (2003), the consistency of virtually all segments (with the exception of C_3) accounted for a significant proportion of variance in at least one analysis. If readers are indeed sensitive to different grain sizes during word recognition (cf. Ziegler & Goswami, 2005), and this sensitivity is modulated by stimulus properties, it is unsurprising that consistency effects can operate at so many different levels of granularity. This position is motivated by Chateau and Jared's (2003) general conclusion that consistency effects for an orthographic segment emerge to the extent that *that* unit constrains pronunciation for ambiguous segments. For example, V_1C_2 effects were stronger in the simultaneous multiple regression (which featured 477 words possessing consistency measures for all orthographic segments), because for such words (e.g., *vertex*), V_1C_2 may constrain the pronunciation of the ambiguous vowel V_1 as well or better than the BOB.

Overview of the present study

The present study examines the influence of major psycholinguistic variables on the word recognition performance of 6115 English monomorphemic multisyllabic words, via hierarchical regression analyses of a large-scale database of behavioral data. Specifically, the targeted behavioral database is the ELP (Balota et al., 2007), which contains lexical and behavioral measures for 40,481 words. The data in the ELP were collected from over 1200 participants across six universities. Examining the effects of variables via multiple regression mitigates many of the limitations associated with factorial experiments, which have played a major role in the field of visual word recognition. Most obviously, factorial designs require variables of interest to be orthogonally manipulated. This is particularly difficult with word stimuli because many lexical properties influence word recognition performance, and these properties are often correlated with each other. For example, shorter words tend to occur more frequently in the language, and this multicollinearity makes the task of selecting items that vary *only* on the variable of interest, while matching for other variables, particularly vexing (Cutler, 1981). The failure to control for extraneous variables has led to a number of controversies in the field. For example, Gernsbacher (1984) demonstrated that the inconsistent interactions between word frequency and a number of variables (bigram frequency, concreteness, polysemy) were due to confounding experiential familiarity with the second variable of interest. More recently, Monaghan and Ellis (2002) have argued that the three-way interaction effect between consistency, frequency, and imageability reported by Strain, Patterson, and Seidenberg (1995) was driven by the failure to control for age of acquisition. Factorial experiments also require continuous

psycholinguistic variables (e.g., frequency, length, consistency) to be converted into categories. There is also evidence that such categorization could yield misleading results, either by decreasing statistical power and reliability (Cohen, 1983) or by inflating the Type II error rate (MacCallum, Zhang, Preacher, & Rucker, 2002). Finally, factorial experiments are more vulnerable to context effects, whereby the characteristics of the words within a list can modulate the effect being studied. For example, Andrews (1997) has argued that some of the inconsistencies in the orthographic N literature are a consequence of variations in the stimulus list environment, and Seidenberg, Waters, Barnes, and Tanenhaus (1984) also found that the effects of regularity depend upon other words in the list that share the same rimes as the experimental stimuli.

The present study extends the large-scale studies by Balota et al. (2004) and Chateau and Jared (2003). Balota et al. examined the effects of surface, lexical, and semantic variables on the pronunciation and lexical decision performance of 2428 monosyllabic words, while Chateau and Jared examined the effects of phonological consistency for various orthographic segments on the pronunciation performance of 1000 six-letter disyllabic words. Here, we explore the effects of a set of variables on visual word recognition for 6115 monomorphemic multisyllabic words in the ELP (see Balota et al., 2007 for a fuller description of this database), breaking new ground on multiple fronts. First, we will investigate lexical processes in both lexical decision and in speeded pronunciation for multisyllabic words; Chateau and Jared only focused on speeded pronunciation. The convergence and divergence across tasks is useful in providing information about the nature of the variables of interest. Second, in the present study, word recognition performance for a greater number and variety of words is considered; Chateau and Jared focused exclusively on six-letter disyllabic words. Third, and most importantly, we explore the effects of novel variables that are optimized for multisyllabic words. For example, Chateau and Jared's analyses provided valuable insights into how the consistency of specific orthographic segments differentially influenced multisyllabic word recognition, but their approach required them to run separate analyses for each orthographic structure. In contrast, the consistency measures we will be using are global and encompass the entire word, effectively allowing us to accommodate all 6115 words within a *single* regression analysis, with their variable lengths, structures, and number of syllables. However, our approach has limited resolution for identifying the impact of *specific* orthographic units. Ultimately, we see both approaches as complementary rather than mutually exclusive, each yielding different but converging pieces of information.

Together with consistency, we also study the effects of traditional (e.g., length, frequency, orthographic and phonological neighborhood size) and newer variables that capture orthographic similarity and semantics for multisyllabic words. The full set of variables used will be described more fully in the next section. In addition to examining the *linear* effects of variables, non-linear (cf. Baayen, Feldman, & Schreuder, 2006) and interactive effects in multisyllabic word recognition will also be explored. A better specification of the basic experimental

Table 1
Means and standard deviations for full set of predictors and dependent variables explored in the item-level regression analyses.

	Mean	SD
Pronunciation RT (*Z*-score)	−0.032	0.460
Pronunciation accuracy	0.905	0.140
LDT RT (*Z*-score)	−0.035	0.425
LDT accuracy	0.774	0.240
Number of syllables	2.373	0.618
Length	6.746	1.566
Word Frequency (rank)	20730	9870
Orthographic N	0.779	1.493
Phonological N	2.334	2.444
Levenshtein orthographic distance	2.640	0.816
Levenshtein phonological distance	4.505	1.764
LOD Neighborhood Frequency	7.008	0.664
S1 feedforward onset consistency	0.934	0.175
S1 feedforward rime consistency	0.470	0.280
S1 feedback onset consistency	0.914	0.191
S1 feedback rime consistency	0.615	0.310
Distance consistency	0.611	0.131
Composite FFO consistency	0.836	0.158
Composite FFR consistency	0.539	0.201
Composite FBO consistency	0.745	0.177
Composite FBR consistency	0.534	0.201
WordNet number of senses	0.509	0.223
Local semantic neighborhood size	2.627	0.553

phenomena associated with multisyllabic words will be critical in defining the constraints that have to be satisfied by nascent and future models of multisyllabic word recognition. We will now turn to a description of our predictor variables.

Predictor variables for the regression analyses

The variables in the analyses were divided into three clusters: surface variables, lexical variables, and semantic variables (see Table 1 for descriptive statistics of predictors and measures). Table 2 presents all the intercorrelations between the predictors and dependent variables being examined.

Surface level

The surface level variables are designed to capture the variance associated with voice key biases and stress patterns. Dichotomous variables were used to code the initial phoneme[1] of each word (1 = presence of feature; 0 = absence of feature) on 13 features: affricative, alveolar, bilabial, dental, fricative, glottal, labiodentals, liquid, nasal, palatal, stop, velar, and voiced (Spieler & Balota, 1997; Treiman et al., 1995). Phonetic bias effects can be attributable either to articulatory (some phonemes take more time to initiate) or acoustic (some phonemes take more time for the voice key to detect) reasons (Kessler et al., 2002). Importantly, as one would expect if these variables are specific to voice key operation, the constellation of 13 variables has been shown to powerfully predict pronunciation (35% of

---

[1] We chose to control for the phonetic properties of only the first phoneme to maintain parity with earlier large-scale studies by Treiman et al. (1995), Chateau and Jared (2002), and Balota et al. (2004), although it should be noted that the second phoneme may also have an effect on response times (Kessler, Treiman, & Mullennix, 2002).

Table 2
Correlations between full set of predictors and dependent variables explored in the item-level regression analyses.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Pronunciation-Z-RT	–	−.68***	.76***	−.67***	.44***	.47***	−.59***	−.29***	−.28***	.55***	.55***	−.29***
2. Pronunciation-Acc		–	−.59***	.75***	−.21***	−.17***	.52***	.13***	.12***	−.24***	−.29***	.08***
3. LDT-Z-RT			–	−.72***	.37***	.39***	−.70***	−.22***	−.20***	.47***	.45***	−.26***
4. LDT-Acc				–	−.13***	−.11***	.70***	.10***	.08***	−.21***	−.21***	.08***
5. Number of syllables					–	.66***	−.17***	−.28***	−.31***	.69***	.75***	−.40***
6. Length						–	−.24***	−.38***	−.37***	.84***	.68***	−.70***
7. Rank composite freq							–	.12***	.11***	−.26***	−.20***	.28***
8. Orthographic N								–	.64***	−.50***	−.45***	.19***
9. Phonological N									–	−.45***	−.50***	.16***
10. L orthographic distance										–	.83***	−.53***
11. L phonological distance											–	−.35***
12. Neighborhood freq												–
13. S1 FF onset consistency												
14. S1 FF rime consistency												
15. S1 FB onset consistency												
16. S1 FB rime consistency												
17. Distance consistency												
18. Compo FFO consistency												
19. Compo FFR consistency												
20. Compo FBO consistency												
21. Compo FBR consistency												
22. WdNet no. senses												
23. Semantic neigh. size												

	13	14	15	16	17	18	19	20	21	22	23
1. Pronunciation-Z-RT	−.15***	.00	−.16***	−.08***	−.22***	−.15***	−.12***	.01	−.13***	−.38***	−.53***
2. Pronunciation-Acc	.05***	.04**	.04***	.05***	.18***	.07***	.11***	−.02	.11***	.28***	.41***
3. LDT-Z-RT	−.04***	.05***	−.04***	−.03*	−.14***	−.05***	−.06***	.04***	−.05***	−.44***	−.63***
4. LDT−Acc	.04***	−.01	.03**	.03*	.08***	.02	.03*	.01	.05***	.36***	.58***
5. Number of syllables	−.01	−.02	.01	.07***	−.33***	−.07***	−.26***	.07***	.00	−.19***	−.19***
6. Length	−.02†	.14***	−.03†	.04**	−.02	−.12***	−.01	−.01	.01	−.13***	−.19***
7. Rank composite freq	.02	−.07***	.01	−.02	−.02†	−.01	−.02	.05***	−.04**	.45***	.81***
8. Orthographic N	.03*	−.05***	.05***	.07***	.15***	.10***	.06***	−.08***	.11***	.18***	.13***
9. Phonological N	.03*	−.03*	.01	.05***	.39***	.12***	.13***	−.22***	.01	.18***	.13***
10. L orthographic distance	−.03*	.11***	−.06***	−.02†	−.10***	−.10***	−.06***	.01	−.08***	−.27***	−.28***
11. L phonological distance	−.06***	.00	−.06***	−.02	−.59***	−.15***	−.22***	.13***	−.08***	−.29***	−.24***
12. Neighborhood freq	.02	−.17***	.00	−.08***	−.20***	−.00	−.09***	.13***	−.06***	.09***	.23***
13. S1 FF onset consistency	–	.00	.53***	.05***	.08***	.46***	−.01	.21***	.07***	.03†	.02
14. S1 FF rime consistency		–	.01	.33***	.18***	−.05***	.67***	.01	.17***	−.05***	−.08***
15. S1 FB onset consistency			–	.06***	.02†	.23***	.01	.46***	.06***	.05***	.04***
16. S1 FB rime consistency				–	.05***	.05***	.21***	−.05**	.65***	.01	−.01
17. Distance consistency					–	.16***	.31***	−.29***	.07***	.15***	.02
18. Compo FFO consistency						–	−.01	.22***	.11***	.02	−.04**
19. Compo FFR consistency							–	.03*	.21***	.03*	−.02
20. Compo FBO consistency								–	−.05**	−.03*	.05**
21. Compo FBR consistency									–	.05***	−.03*
22. WdNet no. senses										–	.45***
23. Semantic neigh. size											–

* $p < .05$.
** $p < .01$.
*** $p < .001$.
† $p < .10$.

the variance), but not lexical decision (1% of variance), performance (Balota et al., 2004). Four dummy variables were also included to capture the stress pattern of a word, i.e., the syllable on which stress falls (Chateau & Jared, 2003); words with stress on the first syllable comprised the reference group.

Lexical level
The lexical variables refer to characteristics that are higher order than phonetic features but lower-level than semantic features. Note that number of phonemes was not included as a predictor due to its very high correlation with length, $r = .836$.

Length (linear). Number of letters.

Length (quadratic). (Number of letters)². New et al. (2006) reported a quadratic U-shaped relationship between length and lexical decision latencies. For short words, length and latencies were negatively correlated, for medium-length words, there was no relationship, and for long words, length and latencies were positively correlated. Quadratic effects of length were examined in both speeded pronunciation and lexical decision performance, after controlling for variables not controlled for by New et al.

Number of syllables. This refers to the number of syllables. In the 6115 multisyllabic words examined, there were

4244 disyllabic words (69.4%), 1495 trisyllabic words (24.4%), 344 quadrasyllabic words (5.6%), 28 pentasyllabic words (.5%), and four hexasyllabic words (.1%).

Orthographic neighborhood size (orthographic N). The number of words that can be obtained by changing a single letter in the target word, while holding the other letters constant (Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, the orthographic neighbors of *cat* include *bat*, *cot*, and *cap*.

Phonological neighborhood size (phonological N). The number of words that can be obtained by changing a single phoneme in the target word, while holding the other phonemes constant (Mulatti, Reynolds, & Besner, 2006; Yates, 2005). For example, the phonological neighbors of *gate* include *hate*, *get*, and *bait*.

Word frequency. The word frequency measure used in this study is the composite rank frequency measure developed by Yap (2007). Yap compared the predictive power of rank vs. log-transformed measures (c.f., Murray & Forster, 2004) for a number of popular word frequency counts, using a large sample of over 26,000 words from the ELP (Balota et al., 2007). The three best frequency measures identified (based on proportion of word recognition variance accounted for) were the HAL norms (Lund & Burgess, 1996), Zeno norms (Zeno, Ivens, Millard, & Duvvuri, 1995), and Google norms (Brants & Franz, 2006). Moreover, rank-transformed counts accounted for more variance than logarithm-transformed (i.e., $\log_{10}(\text{frequency} + 1)$) counts, consistent with Murray and Forster's (2004) argument that recognition times are linearly related to the rank position of a word on a frequency-ordered list. Rank transformation involves sorting all the words in a corpus (in this case, all the words in the ELP) by frequency, and assigning lower ranks to words with higher raw frequency counts. Interestingly, the *mean* of the rank HAL, rank Zeno, and rank Google frequency measures predicted word recognition performance better than any of the individual constituent measures, and it is this composite rank frequency measure that will be employed in the present study.

Levenshtein measures. The Levenshtein measures (Yarkoni, Balota, & Yap, 2008) include Levenshtein orthographic distance (LOD), Levenshtein phonological distance (LPD), and Levenshtein neighborhood frequency (LODNF). These metrics are all based on Levenshtein distance, a measure of string similarity used in information theory and computer science, which is defined as the number of insertions, deletions, and substitutions needed to convert one string of elements (e.g., letters or phonemes) to another. In order to create usable metrics of orthographic and phonological similarity, orthographic and phonological Levenshtein distances were first calculated between every word and every other word in the ELP. LOD and LPD represent the mean orthographic and phonological Levenshtein distances, respectively, from a word to its 20 closest neighbors. LODNF reflects the mean frequency of the 20 closest neighbors. Briefly, the Levenshtein measures circumvent many of the

limitations associated with traditional neighborhood measures (e.g., orthographic N), and in fact are more powerful predictors of performance than traditional variables (see Yarkoni et al.). For example, the utility of orthographic/ phonological N is limited for long words since most long words (e.g., *encyclopedia*) have few or no orthographic/ phonological neighbors, whereas LOD and LPD are useful for words of all lengths (see Fig. 2).

Syllable 1 consistency measures. These consistency measures reflect the feedforward (spelling-to-sound) and feedback (sound-to-spelling) consistency of the onset and rime segments in the *first* syllable of each word. Thus far, we have discussed *feedforward* consistency, which measures the degree to which items with similar spellings have similar pronunciations. Stone, Vanhoy, and Van Orden (1997) have proposed that feedback consistency, which measures the degree to which items with similar pronunciations have similar spellings, also matters. For example, the rime in *plaid* is feedback-inconsistent because /æd/ is usually spelled as *ad* (e.g., *mad*, *cad*, *had*), not *aid*. The feedback consistency literature is somewhat controversial, and some studies find feedback consistency effects in lexical decision (Lacruz & Folk, 2004; Perry, 2003; Stone et al., 1997; Ziegler, Montant, & Jacobs, 1997) and speeded pronunciation (Balota et al., 2004; Lacruz & Folk, 2004; Ziegler et al., 1997) but others do not (Massaro & Jesse, 2005; Peereman, Content, & Bonin, 1998; Ziegler, Petrova, & Ferrand, 2008). It is worth noting that feedback consistency has yet to be explored in multisyllabic word recognition.

All consistency measures in the present study were computed using a corpus based on the 9639 monomorphemic words in the ELP (Balota et al., 2007). To decide what constituted the syllables of a word, the linguistically-defined syllabic boundaries in CELEX (Baayen, Piepenbrock, & van Rijn, 1993) were consulted. In cases where there was a mismatch between syllabic and phonological syllabic boundaries (e.g., for *abacus*, the orthographic parsing is *ab-a-cus* while the phonological parsing is /a-bə-kəs/), the word was re-parsed so that the orthographic parsing was aligned with its phonological counterpart (i.e., *a-ba-cus*). Consistency ranges from 0 to 1, with larger values indicating higher consistency (see Yap, 2007, for more information).

Composite consistency measures and Levenshtein phonological consistency. The composite consistency measures (Yap, 2007) reflect mean consistency across syllabic positions. For example, the composite feedforward rime consistency of the disyllabic word *worship* is the mean feedforward rime consistency of *-or* (Syllable 1) and *-ip* (Syllable 2). Using composite measures allows the full set of words to be included in the regression analyses. If the syllable-specific measures for Syllable 1 and Syllable 2 feedforward rime consistency were included in the regression model, this would limit observations to words which have two or more syllables. Levenshtein phonological consistency (LPC) is the ratio of LOD to LPD (see above), and approaches 1 for consistent words and 0 for inconsistent words. This measure, which is not syllabically defined, should capture consistency across the entire letter string.
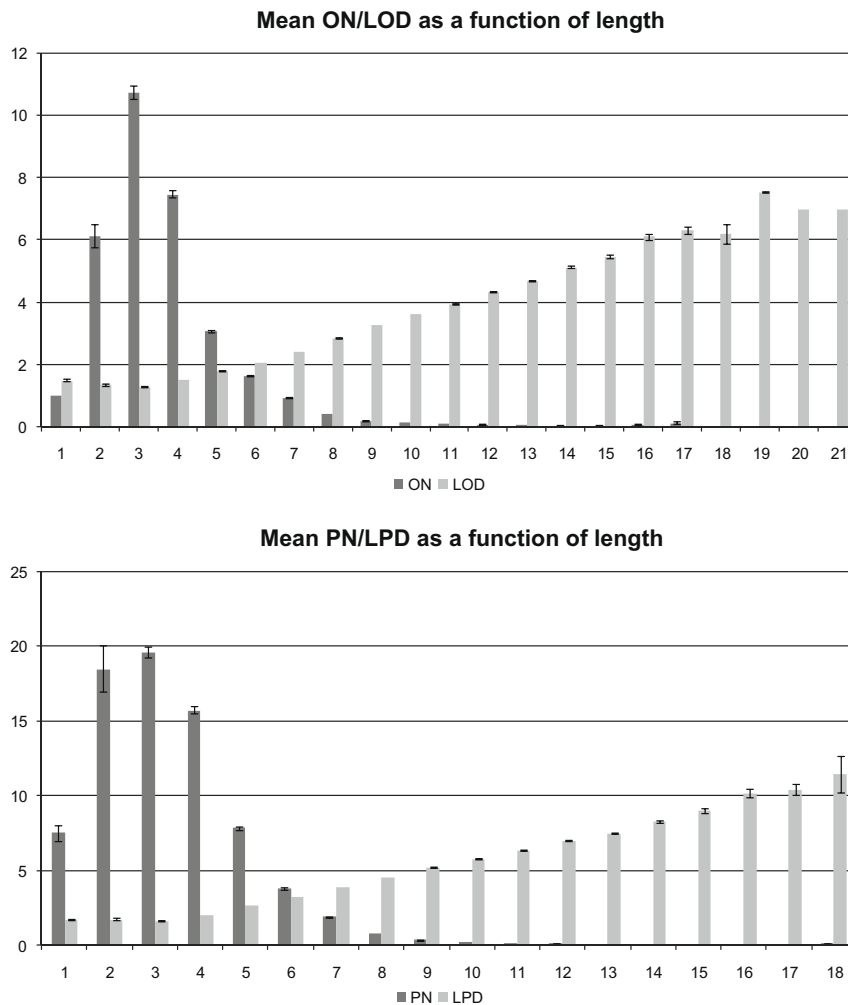
Mean ON/LOD as a function of length

Mean PN/LPD as a function of length

Fig. 2. Mean orthographic N and Levenshtein orthographic distance as a function of length (top panel). Mean phonological N and Levenshtein phonological distance as a function of length (bottom panel). Error bars denote standard errors.

Semantic level
Number of senses. The number of meanings a word possesses in the WordNet database (Miller, 1990); this variable is log-transformed in the present analyses, since the underlying distribution is highly skewed.

Local semantic neighborhood size. Local semantic neighborhood (LSN) size reflects the number of semantic neighbors within a specified radius in high-dimensional semantic space (Durda, Buchanan, & Caron, 2006; http://www.word-mine2.org). Specifically, words with more neighbors within some radius possess denser neighborhoods. This variable is log-transformed in the present analyses.

Item-level regression analyses

Three discrete sets of analyses were conducted. Although there are 9639 monomorphemic words in the ELP, our analyses and subsequent discussion will focus on the 6115 *multisyllabic* words in the monomorphemic

corpus, since multisyllabic words are the focus of this study. In the first section, we will describe item-level regression analyses on the pronunciation and lexical decision performance for 6115 multisyllabic words, which are supplemented by separate analyses for monosyllabic words (n = 3524) and all (i.e., monosyllabic and multisyllabic; n = 9639) words. This is followed by regression analyses exploring theoretically important interactions. In the second section, we present analyses employing measures that are *specific* to multisyllabic words. Note that the analyses in the first section feature measures that are available for *both* mono- and multisyllabic words.

Item-level analyses for all monomorphemic words

An eight-step hierarchical regression analyses was conducted for both pronunciation and lexical decision performance. Phonological onsets were entered in Step 1, stress pattern in Step 2, number of syllables, length, orthographic N, phonological N, and rank composite frequency in Step 3, quadratic length in Step 4, LOD, LPD, and LODNF in Step 5,

first syllable consistency measures in Step 6, LPC and composite consistency measures in Step 7, and semantic measures (number of senses and LSN size) in Step 8. The rationale for this hierarchy was to first enter variables coding onset and prosodic properties, followed by entering established, well-studied lexical variables such as length, word frequency, and neighborhood density. Novel lexical variables (i.e., the Levenshtein-based measures) were then entered to assess their effects above and beyond traditional measures. After lexical variables were controlled for, we then turned to phonological consistency measures based on the first syllable, followed by measures reflecting consistency across the whole letter string. Finally, semantic measures were entered in the final block to see if semantics accounted for a significant amount of variance after various surface, lexical, and phonological consistency measures were controlled for. The eight steps allow us to estimate the regression coefficients for predictors within each step, without the influence of subsequently entered correlated variables. For example, the first syllable (Step 6) and composite (Step 7) consistency measures are obviously correlated, since the composite consistency measures are based in part on the syllable 1 measures. Entering these sets of variables in separate steps allows one to estimate the unique influence of syllable 1 consistency with greater fidelity, and importantly, to assess if composite consistency measures account for additional unique variance after syllable 1 consistency is controlled for.

Tables 3 and 4 present the results of regression analyses after surface and lexical predictors were entered. Tables 5 and 6 present the results of regression analyses when the semantic variables (i.e., LSN and WordNet number of senses) were entered. Note that the regression coefficients reported in the tables reflect the coefficients for variables entered in that particular step, rather than coefficients obtained from entering all variables simultaneously in the model. Analyses of RTs will first be presented, followed by analyses of response accuracy.

Response latencies
Surface level variables (onsets). First, consider the effects of the surface phonological onset variables in speeded pronunciation. Unsurprisingly, the surface variables accounted for more variance in speeded pronunciation than in lexical decision (4.3% vs. 0.3%). However, it is very surprising that the surface variables accounted for so little variance in the multisyllabic dataset. We explored this further by considering the 2428 monosyllabic words used in both this study and the Balota et al. (2004) study. For those words, onsets predicted 35% of the variance in speeded pronunciation performance in the Balota et al. study and 34% of the variance in the present study. To rule out the possibility that onsets are more variable for monosyllabic words, we also compared 3410 monosyllabic and 3410 multisyllabic words which were matched on onsets. Onset characteristics still accounted for 27.8% of the variance in monosyllabic words and only 5.4% in multisyllabic words. It is plausible that computing phonology or programing articulation is considerably less complex for monosyllabic words than for multisyllabic words. Multisyllabic pronun-

ciation implicates additional processes that deal with morphology, stress assignment, vowel reduction, and co-articulation mechanisms. If multisyllabic words indeed exhibit greater variability in pronunciation performance than monosyllabic words, and if one assumes that onset variables can only account for some finite amount of variance, then it follows that onset variables should be able to account for a larger proportion of variance in monosyllabic words, since there is less total variance to explain.

This explanation has to be qualified by the possibility that the large onset effects in monosyllabic speeded pronunciation are spuriously driven by a small subset of words. Specifically, since voice keys have trouble with words beginning with /s/, such words may be responsible for onset variables accounting for so much variance in monosyllabic words. To test this possibility, we examined onset effects in monosyllabic words, with /s/ onset words removed. The results of this analysis were mixed. When /s/ onset words were dropped, the proportion of variance accounted for by onsets dropped from 34% to 14% in the present dataset. However, in the Balota et al. (2004) dataset, this proportion was relatively unchanged, decreasing to 33% from 35%. What could be accounting for these differences? First, the same participants contributed data to *all* words in the Balota et al. study, whereas in the ELP, participants only responded to a subset of the 40,481 words, and different sets of participants contributed data for different words. Second, different voice keys were used for the two studies, and fricatives are particularly sensitive to differences between voice keys. However, it is important to note that even when /s/ onset words are excluded, onsets still account for far *less* variance in multisyllabic words (2.8%) than in monosyllabic words (14%), suggesting that the effect is real. Hence, it is possible that phonetic bias effects, which have been a source of concern in speeded pronunciation experiments (Kessler et al., 2002), may play a more subtle role in multisyllabic word recognition. It is not the case that phonetic biases are irrelevant for multisyllabic words, but it is clear that their effects are overshadowed by other effects, suggesting that they are less likely to spuriously produce effects due to their being confounded with other variables (cf. Kessler, Treiman, & Mullennix, 2008).

Surface level variables (stress). Stress pattern (i.e., the position of the stressed syllable), entered via four dummy-coded variables, also accounted for additional variance reliably after onsets were controlled for (9.3% of variance in speeded pronunciation and 6.4% in lexical decision). Note that words with stress on the first syllable formed the reference group. In both speeded pronunciation and lexical decision, the regression coefficients for the dummy-coded variables indicated that first-syllable-stress words had the shortest RTs, followed by words with second-, third-, and fourth-syllable-stress, respectively. However, these results may simply be due to the statistical covariations between stress pattern and lexical variables. For example, shorter words with fewer syllables (which are recognized faster) are necessarily associated with stress on the earlier syllables. When stress pattern effects were examined *after* controlling for onsets and lexical vari-

Table 3

Standarized RT and accuracy regression coefficients from Steps 1 to 7 of the item-level regression analyses for speeded pronunciation performance for monosyllabic words, multisyllabic words, and all words. The p-value for each R^2 change is represented with asterisks.

Predictor variable	Monosyllabic words (n = 3524)		Multisyllabic words (n = 6115)		All words (n = 9639)	
	RT	Accuracy	RT	Accuracy	RT	Accuracy
Surface variables (onsets)						
R-square	**.280****	**.000**	**.043****	**.000**	**.052****	**.003****
Surface variables (stress)						
			.136**	**.026****	**.188****	**.044****
R-square	NA	NA	$\Delta R^2 = .093$	$\Delta R^2 = .026$	$\Delta R^2 = .136$	$\Delta R^2 = .041$
Standard lexical variables						
Number of syllables	NA	NA	.225***	−.118***	.269***	−.164***
Length (number of letters)	.110***	.013	.121***	.073***	.180***	.055**
Rank composite frequency	−.388***	.448***	−.499***	.506***	−.438***	.489***
Orthographic N	−.231***	.165***	−.072***	.038**	−.082***	.071***
Phonological N	.163***	−.169***	−.042***	.016	.086***	−.090***
	.510**	**.214****	**.537****	**.291****	**.592****	**.302****
R-square	$\Delta R^2 = .230$	$\Delta R^2 = .214$	$\Delta R^2 = .401$	$\Delta R^2 = .265$	$\Delta R^2 = .404$	$\Delta R^2 = .258$
Quadratic length						
Quadratic length	.653***	−.378***	.424***	−.393***	.494***	−.320***
	.519**	**.217****	**.541****	**.295****	**.598****	**.304****
R-square	$\Delta R^2 = .009$	$\Delta R^2 = .003$	$\Delta R^2 = .004$	$\Delta R^2 = .004$	$\Delta R^2 = .006$	$\Delta R^2 = .002$
Distance variables						
L orthographic distance	.022	.060	.142***	.028	.136***	.058*
L phonological distance	.243***	−.243***	.298***	−.297***	.366***	−.385***
LOD neighborhood frequency	.116***	−.124***	.127***	−.137***	.151***	−.165***
	.549**	**.243****	**.588****	**.327****	**.642****	**.337****
R-square	$\Delta R^2 = .030$	$\Delta R^2 = .026$	$\Delta R^2 = .047$	$\Delta R^2 = .032$	$\Delta R^2 = .044$	$\Delta R^2 = .033$
Syllable 1 consistency variables						
Feedforward onset consistency	−.113***	.086***	−.055***	.022†	−.058***	.034**
Feedforward rime consistency	−.053***	.134***	−.023*	.044***	−.029***	.077***
Feedback onset consistency	−.091***	.025	−.043***	.020	−.038***	.015
Feedback rime consistency	−.110***	.062***	−.070***	.037**	−.071***	.041***
	.584**	**.271****	**.600****	**.331****	**.653****	**.345****
R-square	$\Delta R^2 = .035$	$\Delta R^2 = .028$	$\Delta R^2 = .012$	$\Delta R^2 = .004$	$\Delta R^2 = .011$	$\Delta R^2 = .008$
Higher-order consistency variables						
Distance consistency	−.139***	.189***	−.148***	.104***	−.074***	.074**
Composite FF onset consistency	NA	NA	−.048***	.024†	−.045***	.027*
Composite FF rime consistency	NA	NA	−.035**	.041**	−.044**	.068***
Composite FB onset consistency	NA	NA	.053***	−.011	.066***	−.028*
Composite FB rime consistency	NA	NA	−.094***	.114***	−.074***	.103***
	.586**	**.275****	**.612****	**.342****	**.659****	**.351****
R-square	$\Delta R^2 = .002$	$\Delta R^2 = .004$	$\Delta R^2 = .012$	$\Delta R^2 = .011$	$\Delta R^2 = .006$	$\Delta R^2 = .006$

* $p < .05$.
** $p < .01$.
*** $p < .001$.
† $p < .10$.

ables (number of syllables, length, word frequency, orthographic and phonological neighborhood size), stress pattern did not significantly predict RTs in either task. Of course, including a large number of words with varying lengths, structures, and number of syllables may have obscured stress pattern effects in our analysis. However, it is interesting that Chateau and Jared (2003) also noted that stress pattern was a relatively weak predictor of pronunciation latencies in their study. In fact, latencies were reliably faster for first-syllable-stress words only in the analysis when *all* the consistency predictors were entered, i.e., the analysis with 477 CVCCVC words. In contrast, error rates were reliably lower for first-syllable-stress words in all but one analysis, although Chateau and Jared caution that this might be spuriously driven by the relative rarity of second-syllable-stress words in their stimuli. When we restricted our analysis to disyllabic words, we found that first-syllable-stress words were faster, but this advantage was restricted to the speeded pronunciation task. Thus, although there is some evidence for stress pattern RT effects in disyllabic words, the extent to which these effects generalize to longer words remains unclear. On the whole, this seems consistent with reports that stress pattern (Chateau & Jared) and stress typicality (Arciuli & Cupples, 2006) effects are more salient in error rates than in RTs.

Standard lexical variables

Length (number of letters) and number of syllables. As shown in Tables 3 and 4, length and number of syllables were both positively associated with pronunciation and lexical deci-

Table 4

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

Predictor variable	Monosyllabic words ($n = 3524$)		Multisyllabic words ($n = 6115$)		All words ($n = 9639$)	
	RT	Accuracy	RT	Accuracy	RT	Accuracy
Surface variables (onsets)						
R-square	**.005****	**.002**[†]	**.003****	**.002***	**.007**	**.001**
Surface variables (stress)						
			.067**	**.011****	**.094****	**.014****
R-square	NA	NA	$\Delta R^2 = .064$	$\Delta R^2 = .009$	$\Delta R^2 = .087$	$\Delta R^2 = .013$
Standard Lexical Variables						
Number of syllables	NA	NA	.164***	−.076***	.213***	−.120***
Length (number of letters)	−.098***	.173***	.081***	.134***	.068***	.172***
Rank composite frequency	−.726***	.718***	−.646***	.708***	−.644***	.731***
Orthographic N	−.145***	.116***	−.061***	.050***	−.055***	.063***
Phonological N	.071***	−.110***	−.002	−.013	.069***	−.095***
	.533**	**.496****	**.573****	**.494****	**.619****	**.504****
R-square	$\Delta R^2 = .528$	$\Delta R^2 = .494$	$\Delta R^2 = .506$	$\Delta R^2 = .483$	$\Delta R^2 = .525$	$\Delta R^2 = .490$
Quadratic length						
Quadratic length	.592***	−.328***	.593***	−.369***	.648***	−.403***
	.541**	**.498****	**.580****	**.497****	**.630****	**.509****
R-square	$\Delta R^2 = .008$	$\Delta R^2 = .002$	$\Delta R^2 = .007$	$\Delta R^2 = .003$	$\Delta R^2 = .011$	$\Delta R^2 = .005$
Distance variables						
L orthographic distance	.054	−.036	.186***	−.155***	.179***	−.149***
L phonological distance	.103***	−.094***	.187***	−.147***	.200***	−.166***
LOD neighborhood frequency	.113***	−.160***	.097***	−.135***	.116***	−.171***
	.551**	**.512****	**.610****	**.522****	**.653****	**.530****
R-square	$\Delta R^2 = .010$	$\Delta R^2 = .014$	$\Delta R^2 = .030$	$\Delta R^2 = .025$	$\Delta R^2 = .023$	$\Delta R^2 = .021$
Syllable 1 consistency variables						
Feedforward onset consistency	−.029*	.020	−.010	.024*	−.014[†]	.024**
Feedforward rime consistency	−.030*	.050***	.017[†]	.010	−.008	.039***
Feedback onset consistency	−.012	−.005	−.019[†]	.005	−.015*	.003
Feedback rime consistency	−.022[†]	.015	−.037***	.021*	−.027***	.017*
	.553**	**.514****	**.612****	**.523****	**.654****	**.532****
R-square	$\Delta R^2 = .002$	$\Delta R^2 = .002$	$\Delta R^2 = .002$	$\Delta R^2 = .001$	$\Delta R^2 = .001$	$\Delta R^2 = .002$
Higher-order consistency variables						
Distance consistency	−.057	.091*	−.101***	.072**	−.094***	.085***
Composite FF onset consistency	NA	NA	−.008	−.003	−.010	−.008
Composite FF rime consistency	NA	NA	−.013	.008	−.017	.018
Composite FB onset consistency	NA	NA	.054***	−.003	.043***	.003
Composite FB rime consistency	NA	NA	−.037**	.048***	−.029**	.042**
	.553	**.515***	**.616****	**.525****	**.656****	**.533****
R-square	$\Delta R^2 = .000$	$\Delta R^2 = .001$	$\Delta R^2 = .004$	$\Delta R^2 = .002$	$\Delta R^2 = .002$	$\Delta R^2 = .001$

[*] $p < .05$.
[**] $p < .01$.
[***] $p < .001$.
[†] $p < .10$.

sion latencies. Unsurprisingly, as the length increased, recognition times became slower. The quadratic effect of length was also significant in both tasks, although it accounted for more unique additional variance in lexical decision (R^2 change = 0.7%) than in pronunciation (R^2 change = 0.4%) performance. More interestingly, number of syllables was positively correlated with response times. In order to test this more rigorously, a secondary analysis was conducted where the unique effect of number of syllables was assessed in pronunciation and lexical decision performance, after *all* the surface, lexical, and consistency variables were controlled for, along with an additional measure of length which reflects the number of phonemes. The results were clear. In a large database of monomorphemic multisyllabic words, controlling for all relevant variables, number of syllables was strongly and positively correlated with pronunciation times, $\beta = .077$, $p < .001$, and lexical decision times, $\beta = .049$, $p < .001$.

Orthographic and phonological N. Orthographic N was negatively correlated with RTs in speeded pronunciation and lexical decision, indicating that words with many orthographic neighbors were recognized faster, even after phonological N was controlled for. Furthermore, the observation that facilitatory effects of orthographic N were stronger in pronunciation than in lexical decision is consistent with the notion that these effects could reflect the sublexical orthography-to-phonology procedure, which is emphasized to a greater extent in pronunciation than in lexical decision (see Andrews, 1997, for a discussion of this perspective). In contrast, phonological N facilitated pronunciation, but not lexical decision, performance, which

Table 5
Standardized RT and accuracy regression coefficients for Step 8 (semantic predictors) of the item-level regression analyses for speeded pronunciation performance for monosyllabic words, multisyllabic words, and all words. The *p*-value for each R^2 change is represented with asterisks.

Predictor variable	Monosyllabic words (n = 3226)		Multisyllabic words (n = 5176)		All words (n = 8402)	
	RT	Accuracy	RT	Accuracy	RT	Accuracy
Surface Variables						
R-square	**.293**[***]	**.000**	**.128**[***]	**.019**[***]	**.182**[***]	**.036**[***]
All lexical variables						
R-square	**.593**[***]	**.274**[***]	**.618**[***]	**.347**[***]	**.663**[***]	**.355**[***]
Semantic variables						
Local semantic neighborhood size	.007	−.041[*]	−.042[**]	−.067[***]	−.036[***]	−.035[**]
WordNet number of senses	−.064[***]	.011	−.024[*]	−.022[†]	−.006	−.040[***]
R-square	**.596**[**]	**.274**[†]	**.619**[**]	**.348**[***]	**.663**[***]	**.357**[***]

[*] *p* < .05.
[**] *p* < .01.
[***] *p* < .001.
[†] *p* < .10.

Table 6
Standardized RT and accuracy regression coefficients for Step 8 (semantic predictors) of the item-level regression analyses for lexical decision performance for monosyllabic words, multisyllabic words, and all words. The *p*-value for each R^2 change is represented with asterisks.

Predictor variable	Monosyllabic words (n = 3226)		Multisyllabic words (n = 5176)		All words (n = 8402)	
	RT	Accuracy	RT	Accuracy	RT	Accuracy
Surface variables						
R-square	**.004**[**]	**.003**	**.056**[***]	**.009**[***]	**.085**[***]	**.012**[***]
All lexical variables						
R-square	**.581**[***]	**.520**[***]	**.625**[***]	**.525**[***]	**.668**[***]	**.536**[***]
Semantic variables						
Local semantic neighborhood size	.019	.008	−.075[***]	.023	−.033[***]	.019[†]
WordNet number of senses	−.122[***]	.052[***]	−.066[***]	−.005	−.083[***]	.014
R-square	**.591**[***]	**.522**[**]	**.630**[***]	**.525**	**.673**[***]	**.536**[†]

[**] *p* < .01.
[***] *p* < .001.
[†] *p* < .10.

is inconsistent with studies showing facilitation in lexical decision (Yates, 2005; Yates, Locker, & Simpson, 2004). However, for monosyllabic words in both tasks, when orthographic and phonological N were entered together, one observes facilitation for orthographic N but *inhibition* for phonological N. These intriguing findings will be discussed further in the General Discussion.

Word frequency. Composite rank frequency was negatively correlated with pronunciation and lexical decision latencies, with shorter latencies for more frequent words. The regression analyses (see Tables 3 and 4) also indicate that the predictive power of word frequency was larger in lexical decision than in speeded pronunciation performance, consistent with what Balota et al. (2004) found. Word-frequency effects may be exaggerated in lexical decision because of the task's emphasis on frequency-based information for discriminating between familiar words and unfamiliar non-words (cf. Balota & Chumbley, 1984).

Levenshtein orthographic distance (LOD), Levenshtein phonological distance (LPD), and Levenshtein neighborhood frequency (LODNF). Turning to the distance measures, LOD, LPD, and LODNF produced reliable effects in both pronun-

ciation and lexical decision, such that all were positively correlated with RTs. That is, words that are orthographically and phonologically more distinct were recognized more slowly. Similarly, words which possess neighbors with higher word frequencies were recognized more slowly. It is very clear that these new measures of orthographic and phonological similarity are powerful predictors compared to traditional measures of orthographic and phonological neighborhoods (see detailed analyses in Yarkoni et al., 2008). This is likely due to the greater utility of these measures for longer words, as discussed earlier. When orthographic and phonological neighborhood were entered *after* the Levenshtein measures, orthographic neighborhood size no longer accounted for unique variance, and the regression coefficients for phonological neighborhood size reverse in speeded pronunciation, β = .029, *p* = .012, and lexical decision, β = .046, *p* < .001, suggesting suppression.

Syllable 1 phonological consistency. Syllable 1 consistency measures accounted for more incremental variance in speeded pronunciation (1.2%) than in lexical decision (0.2%). For Syllable 1 phonological consistency measures, feedforward and feedback consistency for both onsets

and rimes, were all negatively correlated with pronunciation RTs. Increasing the consistency of the spelling-sound mapping shortened pronunciation times; this operated in both directions, from spelling-to-sound, and from sound-to-spelling. In lexical decision, only *feedback* rime consistency produced a reliable effect, with consistent words eliciting faster lexical decision RTs (Stone et al., 1997). Interestingly, in lexical decision, facilitatory effects of feed-forward consistency were observed for monosyllabic, but not multisyllabic, words.

Higher order phonological consistency. Reliable effects of higher-order consistency were observed in both tasks, although consistency again accounted for more incremental variance in pronunciation. These higher-order consistency variables include: (1) composite consistency measures that average consistency across syllables, and (2) a consistency measure based on Levenshtein distances. The former reflect consistency mappings beyond the initial syllable, while the latter reflects consistency mappings that transcend syllabic boundaries. As mentioned earlier, Levenshtein phonological consistency is not syllabically defined, and hence should be sensitive to grain sizes both smaller and larger than syllables. For speeded pronunciation, effects of distance consistency and all four composite consistency measures were significant; specifically, with the only exception being feedback onset consistency,[2] more consistent items were associated with faster latencies. For lexical decision performance, Levenshtein distance consistency was also significant, with consistent items producing faster latencies. However, for the composite measures, only feedback onset and feedback rime consistency effects were significant in lexical decision performance; latencies to feedback rime consistent words were faster while latencies to feedback onset consistent words were slower. Collectively, these results indicate that consistency effects, especially in pronunciation, operate beyond the first syllable. Hence, even though syllabically-defined onset and rime consistency are important aspects of phonological consistency in multisyllabic words, it is clear that consistency mappings for non-syllabic grain sizes are also important.

Semantic variables

Finally, it is clear that semantic measures were able to account for some unique variance in pronunciation and lexical decision (see Tables 5 and 6) after other relevant variables were controlled for. Generally, semantic variables accounted for more incremental variance in lexical deci-

sion (0.5%) than in pronunciation (0.1%) performance, which is consistent with lexical decision's emphasis on semantic information for word–non-word discrimination. Local semantic neighborhood size was negatively correlated with RTs in both pronunciation and lexical decision. Specifically, words associated with denser semantic neighborhoods (i.e., more neighbors within a certain radius) were recognized more quickly. WordNet number of senses was negatively correlated with RTs in lexical decision and pronunciation; words with more meanings were recognized more quickly. Generally speaking, there are reliable (albeit subtle) effects of semantic variables in word recognition performance, even after a very substantial proportion of variance in the dependent measures has already been accounted for.

Response accuracy
Surface level variables. Onsets accounted for virtually no variance in speeded pronunciation and lexical decision accuracy (see Tables 3 and 4). This is consistent with the notion that onset coding primarily influence RTs, rather than accuracy, due to temporal biases in voice key sensitivity. In both tasks, stress pattern also accounted for relatively little variance in accuracy. However, even after onsets and lexical covariates were controlled for, stress pattern accounted for unique variance in both tasks. In speeded pronunciation, first-syllable-stress words had reliably more accurate responses than words with second, third, and fourth-syllable-stress. In lexical decision, the accuracy of first-syllable-stress words was significantly higher than that of third- and fourth-syllable-stress words. As discussed earlier, this is in line with the general finding that stress effects are more readily observed in error rates than in RTs (Arciuli & Cupples, 2006; Chateau & Jared, 2003).

Lexical and semantic variables. For the standard lexical variables, the pattern in accuracy broadly mirrored the pattern in RTs (see Tables 3 and 4). There were two notable exceptions to this trend. First, in both speeded pronunciation and lexical decision, length had inhibitory effects in RTs but facilitatory effects in accuracy. To explore this discrepancy, we first examined the zero-order correlations between length and accuracy on the two tasks. Length was weakly and negatively correlated with accuracy on both speeded pronunciation ($r = -.171$) and lexical decision ($r = -.107$); at the zero order, length has weak inhibitory effects on accuracy. We next computed standardized residuals for accuracy for both tasks, after partialling out surface variables, frequency, number of syllables, and neighborhood size. We then used a median split to classify words as short or long, before computing mean standardized residuals as a function of length. Here, we found that short words were associated with lower accuracy on both tasks, consistent with the regression analysis. We do not have a good explanation for this finding, but suggest that it should be interpreted with caution. Zero-order length effects in accuracy were relatively small, and given length's moderate to large correlations with other lexical variables, it is likely that the reversal in the length coefficient could be due to a suppressor relationship between length and the other lexical variables.

---

[2] The observation of inhibitory composite feedback onset (FBO) consistency effects in both tasks (i.e., slower latencies for words with greater feedback onset consistency) is anomalous and perplexing. Clearly, every other consistency measure in the regression model is producing facilitatory effects. Follow-up analyses indicate that this inhibitory effect is only observed with post syllable 1 onsets. That is, in multisyllabic words, syllable 1 FBO consistency produces facilitatory effects, but the FBO consistency of later syllables (i.e., syllable 2 for disyllabic words, and syllables 2 and 3 for trisyllabic words) produce inhibitory effects. These trends are also present in zero-order correlations, suggesting that the inhibitory pattern is not simply due to composite FBO consistency being entered with other predictors in the regression model.

Similarly, regression coefficients for the two semantic variables were in the wrong direction for speeded pronunciation accuracy (but not for lexical decision). Specifically, words with more senses and denser semantic neighborhoods produced more errors. Since this did not occur in lexical decision, it appears to be a task-specific effect. That could be explained as greater competition between semantic neighbors taking place at the level of phonology or articulation, which lowers overall pronunciation accuracy. It is strange that this effect should be observed in accuracy but not in RTs. Clearly, more work needs to be done to address this puzzle.

Theoretically motivated interactions

The foregoing analyses dealt with the *main* effects of different variables on word recognition performance. We also selected and tested a number of interactions: (1) the length × frequency interaction, (2) the orthographic N × frequency interaction, (3) the consistency × frequency interaction, (4) the number of syllables × frequency interaction, and (5) the consistency × frequency × semantics interaction. So far, most of these interactions have been explored primarily with monosyllabic words, and hence, it is important to determine if they are present in multisyllabic words. Obviously, an almost limitless number of interactions could potentially be tested, but these interactions were targeted because of their theoretical importance. The theoretical rationale for exploring each interaction will be provided before the corresponding analyses.

Regression interactions were tested using the technique advocated by Cohen, Cohen, West, and Aiken (2003). Specifically, the variables of interest (and other control variables) were first entered in the regression model, followed by the interaction term in the following step, and the *R*-square change between the two regression models (without and with the interaction term) was then measured. Essentially, this method tests the interaction while controlling for the main effects of the variables, along with other potentially confounding variables. Note that this method uses the full regression model, and statistical power is maximized because continuous variables are not reduced to categories. For reliable interactions, a plot of the interaction is made by computing the slope of one variable at different levels of the other variable. Unless stated otherwise, the same standard lexical variables (i.e., surface variables, length, number of syllables, word frequency, orthographic N, phonological N) were entered as control variables in all these analyses.

Length × frequency interaction. Weekes (1997) reported that length effects were stronger for low-frequency, compared to high-frequency, words. This finding has been taken as critical support for the dual route cascaded (DRC) model of word recognition (Coltheart et al., 2001), because the recognition of low-frequency words is more likely to reflect the serial, sublexical procedure while the recognition of high-frequency words is assumed to reflect the parallel lexical procedure. Fig. 3 presents the length × frequency interaction for the 6115 multisyllabic words. The length × frequency interaction was reliable in pronunciation, $\beta = -.105$, $p < .001$, and approached significance in lexical decision, $\beta = -.014$, $p = .092$; adding phoneme length as a covariate did not modulate the pattern. In both tasks, as word frequency increased, length effects became smaller (see Fig. 3), and the interaction is much larger in pronunciation than in lexical decision performance. While these results are certainly consistent with the serial mechanism in the dual route framework, it is important to point out that a serial procedure is not the only way to produce length effects. For example, length effects may also be generated through *dispersion* (i.e., longer words are associated with less frequent and more difficult spelling-sound correspondences) or peripheral visual and articulatory processes (see Perry, Ziegler, & Zorzi, 2007, for more discussion). Hence, other perspectives may in principle be able to accommodate the interaction.
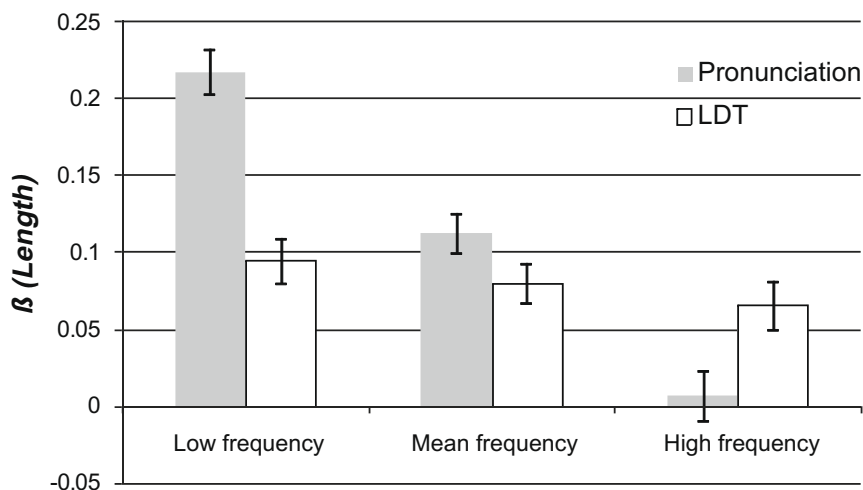

Fig. 3. Word frequency × length interaction. The bars represent the standardized regression coefficient for length as a function of high-, medium-, and low-frequency words, for both pronunciation and lexical decision. Error bars denote standard errors.
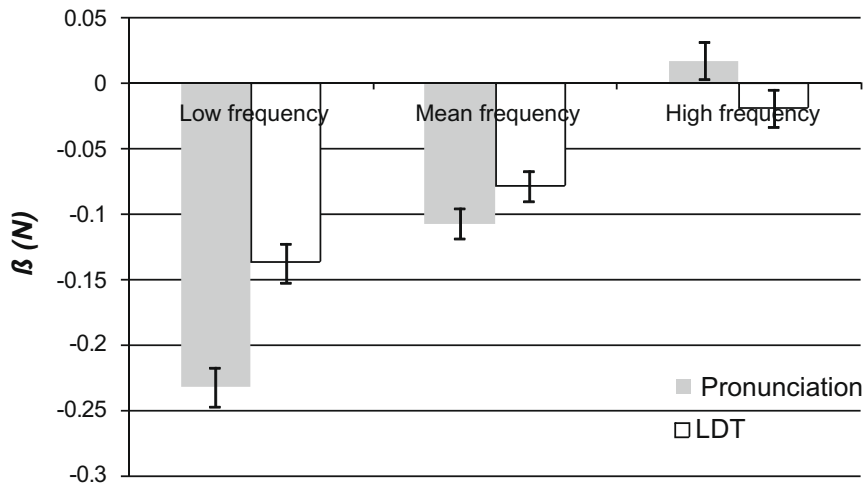
Fig. 4. Word frequency × orthographic N interaction. The bars represent the standardized regression coefficient for orthographic N as a function of high-, medium-, and low-frequency words, for both pronunciation and lexical decision. Error bars denote standard errors.

Orthographic N × frequency interaction. Andrews (1989, 1992) demonstrated that the facilitatory effects of orthographic N (i.e., faster latencies when there are more neighbors) were stronger for low-frequency words than for high-frequency words, in both speeded pronunciation and lexical decision. Using the interactive activation framework (McClelland & Rumelhart, 1981), Andrews (1989) argued that presenting a word activates the word and its orthographic neighbors, and these partially activated units activate sublexical units (e.g., letters) and are in turn activated by them. Words with many neighbors benefit more from this reciprocal activation because more units are involved. More importantly, she speculated that high-frequency words show smaller orthographic N effects because they achieve recognition threshold so quickly that they are less affected by the reverberations between lexical and sublexical units. Alternatively, if one assumes that facilitatory effects of orthographic N reflect sublexical processes (cf. Andrews, 1997), then it follows that facilitatory orthographic neighborhood effects should be more pronounced for lower frequency words (since these words are more likely to be influenced by the sublexical pathway). Although the interaction has been observed across different studies (see also Sears, Hino, & Lupker, 1995), it is interesting that this is one of the few effects that both the DRC (Coltheart et al., 2001) and the CDP+ (Perry et al., 2007) models have difficulty simulating. The present results clearly extend the interaction observed with monosyllabic words to multisyllabic words. Specifically, the orthographic N × frequency interaction was significant in both pronunciation, $\beta = .125$, $p < .001$, and lexical decision, $\beta = .059$, $p < .001$, although the effect was stronger in pronunciation. In both tasks, the facilitatory effects of orthographic neighborhood effects were strongest for low-frequency and weakest for high-frequency words (see Fig. 4).

Consistency × frequency interaction. The regularity × frequency interaction, where low-frequency words show lar-

ger regularity effects than high-frequency words (Andrews, 1982; Seidenberg et al., 1984) is one of the benchmark effects in the word recognition literature. This interaction was initially viewed as support for the dual route model, although it was later shown that the connectionist perspective (Seidenberg & McClelland, 1989) could also handle the interaction. Subsequent work pitting regularity against consistency (e.g., Cortese & Simpson, 2000; Jared, 2002) suggests that consistency, compared to regularity, better captures the mapping between spelling and sound. Interestingly, Jared examined consistency and word-frequency effects in a tightly controlled factorial study and observed a weak but reliable interaction (by subjects) between the two variables, with larger consistency effects for low-frequency words. However, it is important to point out that the interaction was *not* significant by items, suggesting that it is less robust than the literature suggests and may be driven by a confound between word frequency and the neighborhood characteristics of words. Specifically, there may be larger consistency effects for low-frequency words because low-frequency words are more likely to possess neighborhoods where the summed frequency of friends is low relative to the summed frequency of enemies. To test the interaction, the Levenshtein distance-based consistency measure,[3] which works as a global measure of consistency across the letter string, was used. The consistency × frequency interaction was significant in both pronunciation, $\beta = .082$, $p < .001$, and in lexical decision, $\beta = .047$, $p < .001$; the interaction was also stronger in pronunciation (see Fig. 5). Consistency effects (i.e., faster latencies for consistent words) were strongest for low-frequency words, decreasing in magnitude as word frequency increased; this occurred in both tasks but

---

[3] We also tested the word frequency × consistency interaction using composite feedforward rime consistency. The same interactive pattern was observed in speeded pronunciation ($p = .001$) but not in lexical decision ($p = .70$), suggesting that the Levenshtein-based consistency measure may be more sensitive than the composite feedforward rime measure.
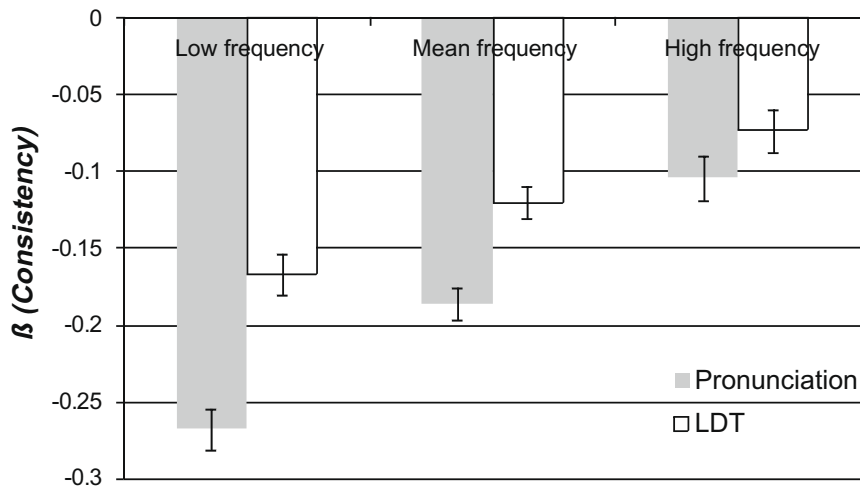
Fig. 5. Word frequency × consistency interaction. The bars represent the standardized regression coefficient for consistency as a function of high-, medium-, and low-frequency words, for both pronunciation and lexical decision. Error bars denote standard errors.

was more salient in pronunciation. Interestingly, although consistency effects were weaker for high-frequency words, they were still reliable in both speeded pronunciation ($p < .001$) and lexical decision ($p < .001$), which meshes well with Jared's (1997) and Jared's (2002) finding of consistency effects for high-frequency words.

Number of syllables × frequency. Jared and Seidenberg (1990, Experiment 3) demonstrated that number of syllables was related to speeded pronunciation latencies, but only for low-frequency words. Since number of syllables effects was absent for words with familiar orthographic patterns (i.e., high-frequency words), Jared and Seidenberg suggested that it is unlikely that there is an explicit syllabification procedure for *all* words. This interaction has been replicated in French, in both speeded pronunciation (Ferrand, 2000) and lexical decision (Ferrand & New, 2003,

Experiment 2). Fig. 6 presents the number of syllables × frequency interaction for the 6115 multisyllabic words. The number of syllables × frequency interaction was reliable in speeded pronunciation, $\beta = -.118$, $p < .001$, and in lexical decision, $\beta = -.025$, $p = .004$. In both tasks, as word frequency increased, number of syllables effects became smaller, and the interaction was substantially larger in pronunciation than in lexical decision performance. The larger number of syllables effect observed for low-frequency words is consistent with previous studies, but it is noteworthy that even high-frequency words show reliable (albeit smaller) effects of number of syllables (ps < .001 for both speeded pronunciation and lexical decision).

Consistency × frequency × semantics. Strain et al. (1995) reported an intriguing interaction between spelling-to-sound consistency, word frequency, and imageability;
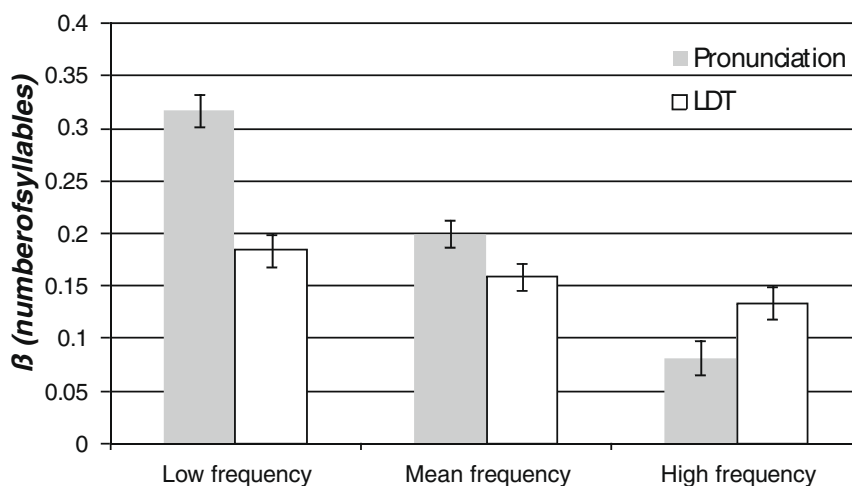

Fig. 6. Word frequency × number of syllables interaction. The bars represent the standardized regression coefficient for number of syllables as a function of high-, medium-, and low-frequency words, for both pronunciation and lexical decision. Error bars denote standard errors.

low-frequency words with inconsistent spelling-to-sound mappings produced the largest imageability effects. The Strain et al. finding was viewed as consistent with Plaut et al.'s (1996) triangle model, which predicts that semantic representations should exert stronger effects when the orthography-to-phonology pathway is noisy, as it is when low-frequency, inconsistent words are processed. This three-way interaction was tested in the present data-set, with Levenshtein distance consistency used as a proxy for consistency. Unfortunately, imageability norms are not available for most of the multisyllabic words. Since it is unclear if local semantic neighborhood size or WordNet number of senses function as better proxies for semantics, the three-way interaction was tested independently with each of the two semantic measures. For both tasks, the three-way interaction was not significant whether local semantic neighborhood size (speeded pronunciation, $p = .99$; lexical decision, $p = .48$) or WordNet number of senses (speeded pronunciation, $p = .21$; lexical decision, $p = .20$) was used. This is in line with Balota et al.'s (2004) conclusion that interactive effects of meaning-level variables are very small, after they failed to detect the consistency $\times$ frequency $\times$ imageability interaction in their database of 2428 monosyllabic words (compared to over 5000 multisyllabic words in the present analyses). Of course, our results depend on the measures of consistency and semantics we used, and it is possible that the interaction may emerge when alternative measures are used.

Analyses specific to multisyllabic words
Context-sensitive phonological consistency. Thus far, phonological consistency has been operationalized in a fairly simplistic manner, either defined narrowly at the level of onsets and rimes *within* individual syllables, or at a more holistic level by examining the discrepancy between Levenshtein orthographic and phonological distances. The present results indicate that syllabically-defined consistency measures account for unique variance, supporting the idea that syllables play a role in visual word recognition. Of course, it is not our claim that syllables are the *only* sublexical unit that mediates lexical access. We have discussed how consistency probably also operates at the level of larger, higher-order units that span syllabic boundaries. For example, BOB effects (Chateau & Jared, 2003) are evidence that readers are able to take advantage of extrasyllabic contextual cues for determining the pronunciation of a rime (e.g., using *d* in *mea-dow* to constrain the pronunciation of *ea*). The extrasyllabic hypothesis can be tested by assessing if contextual consistency accounts for word recognition performance above and beyond consistency defined *within* the syllable. In the following analysis, we tested four new consistency measures (B. Kessler, personal communication, September 2, 2006; see Yap, 2007, for further details) that take into account contextual information *outside* the syllable. Specifically, these measures are: (1) Syllable 1 rime feedforward consistency, taking syllable 2 onset spelling into account; (2) Syllable 1 rime feedforward consistency, taking syllable 2 onset phonology into account; (3) Syllable 1 rime feedback consistency, taking syllable 2 onset spelling into account; and (4) Syllable 1 rime feedback consistency, taking syllable 2 onset phonol-

ogy into account. These are conditional probability measures that consider the identity of an additional unit in a separate syllable. For example, the conditional consistency of the correspondence $er \rightarrow /@'/$ in syllable 1, given the spelling onset *d* in the second syllable, is defined as the consistency of $er \rightarrow /@'/$, examining *only* words with the onset *d* in the second syllable (Kessler & Treiman, 2001). To make this more concrete, consider the rime *us*, which in English, typically maps onto /Vs/ when it occurs in the first syllable (e.g., *custard*, *gusto*). It only rarely maps onto /Vz/ (e.g., *muslin*) or /Uz/ (e.g., *muslim*); these two words have low syllable 1 feedforward rime consistency. However, in disyllabic monomorphemic words, the rime *us* in syllable 1 is followed by the letter *l* for *only* two words, *mus-lin* and *mus-lim*. Hence, if one considers only words where *us* is followed by *l* in the second syllable, the feedforward rime consistency of *us* goes up to .5 for each of these two words, which is somewhat higher than their unconditional consistency.

These measures were entered into a regression model after syllabically-defined consistency measures and other lexical variables were entered (see Table 7). In order to ensure comparability with Chateau and Jared's (2003) disyllabic study, only disyllabic words were included. Even after syllable 1 consistency measures and composite consistency measures were entered into the model, the contextual consistency measures still accounted for an additional 0.4% ($p < .001$) of variance in speeded pronunciation and an additional 0.1% ($p = .009$) of variance in lexical decision. Unsurprisingly, contextual consistency accounted for more variance in pronunciation than in lexical decision performance. In both tasks, syllable 1 feedback rime consistency given syllable 2 onset phonology was facilitatory, suggesting that sound-to-spelling consistency effects for the rime in the *first* syllable is modulated by the pronunciation of the onset in the *second* syllable. The syllable 1 feedforward rime consistency | syllable 2 onset phonology effect was also borderline significant ($p = .05$) in speeded pronunciation, indicating that the constraining influence of the second syllable onset applies to first syllable spelling-to-sound consistency as well. Collectively, even though these effects are small compared to the syllabic consistency effects, they indicate that readers are also able to take advantage of extrasyllabic contextual information to help them constrain the pronunciation of the rime in the first syllable, in both pronunciation and lexical decision. These results are also broadly compatible with Chateau and Jared's (2003) BOB effects.

General discussion

This present study examined the influence of major psycholinguistic variables on the word recognition performance of 6115 monomorphemic multisyllabic English words. Extant theories and models in visual word recognition have been overwhelmingly informed by the study of monosyllabic words, and researchers in the word recognition domain have consistently pointed towards multisyllabic words as *the* important next step.

Table 7
Standarized RT regression coefficients for contextual consistency measures (Step 6) in speeded pronunciation and lexical decision performance.

	R-square p	P (R-sq change)
Speeded pronunciation (n = 4243)		
Step 1: phonological onsets	.090	<.001
Step 2: stress	.102	<.001
Step 3: lexical variables	.507	<.001
Step 4: syllable 1 phonological consistency	.523	<.001
Step 5: composite consistency	.534	<.001
Step 6: contextual consistency	.538	<.001
	beta	p-value
FF S1 rime consistency| S2 onset spelling	−.017	ns
FF S1 rime consistency| S2 onset phonology	−.041	.050
FB S1 rime consistency| S2 onset spelling	.005	ns
FB S1 rime consistency| S2 onset phonology	−.080	<.001
Lexical decision (n = 4243)		
Step 1: phonological Onsets	.008	<.001
Step 2: stress	.011	.002
Step 3: lexical variables	.560	<.001
Step 4: syllable 1 phonological consistency	.561	.001
Step 5: composite consistency	.564	<.001
Step 6: contextual consistency	.565	.004
	beta	p-value
FF S1 rime consistency| S2 onset spelling	.021	ns
FF S1 rime consistency| S2 onset phonology	−.034	.094
FB S1 rime consistency| S2 onset spelling	.058	.010
FB S1 rime consistency| S2 onset phonology	−.080	<.001

The approach we have taken can be extended in a number of different directions. The most obvious and natural extension is to the reading of multisyllabic words. (Plaut et al., 1996, p. 106).

All of the current computational models of reading English, including the DRC model, are restricted to the processing of monosyllabic stimuli. The problems encountered when items of more than one syllable are considered are numerous… [and] have been underrepresented in the psycholinguistic literature and virtually ignored in modeling reading. (Coltheart et al., 2001, p. 249).

Finally, CDP+ is limited to monosyllabic words, as are the other computational models we have discussed. This reflects the fact that most of the empirical evidence on written word naming comes from research conducted with monosyllabic words. Although most of the words people read are monosyllabic according to a token count, the majority of the words in the lexicon are polysyllabic according to a type count. (Perry et al., 2007, p. 304).

The results reported in this paper represent important new constraints for present and emerging models of multisyllabic visual word recognition.

What are the specific multisyllabic effects that visual word recognition models need to accommodate?

One of the interesting observations that emerged from the present analyses is that by and large, the processing of multisyllabic words does *not* appear to be radically different from the way monosyllabic words are processed. For example, if one were to compare effects in monosyllabic words against effects in multisyllabic words (see Tables 3 and 4), there is a high degree of concordance with respect to how the classic lexical variables (e.g., word frequency, orthographic N, phonological consistency) behave. On the whole, this is encouraging, and confirms that experimental work based on monosyllabic words is *not* simply idiosyncratic to monosyllabic words. However, it is also important to point out that studying multisyllabic words yielded a number of important observations that could not have been obtained simply by studying monosyllabic words. These effects will now be highlighted in the following sections.

The role of onsets

One of the surprising findings in this study is that phonological onsets account for relatively less variance in multisyllabic than in monosyllabic pronunciation performance. This might be due to multisyllabic pronunciation being necessarily more complex, involving additional mechanisms for stress assignment and vowel reduction. If onsets can only account for a finite amount of variance, and there is more total variance to explain in multisyllabic word performance, it follows that onsets will account for a smaller proportion of total variance when such items are examined. The more important point here is that one must be cautious about generalizing the size of an effect from monosyllabic words to larger classes of words.

The role of the syllable

As discussed in the Introduction, some researchers in visual word recognition are skeptical that the syllable plays an important role in lexical access, given that not all English words possess clear syllabic boundaries. Our results are consistent with Chateau and Jared's (2003) view that there are multiple codes mediating lexical access and output processes, and the syllable is one of those codes. Two pieces of evidence are relevant here. First, pronunciation and lexical decision latencies were positively correlated with the number of syllables in a word, even after important length covariates such as length and phonemes, along with other covariates like word frequency, neighborhood size, and phonological consistency were controlled for. The consistency control variables are particularly important because Jared and Seidenberg (1990) have argued that number of syllables effects are artifactually driven by number of syllables being confounded with phonological consistency. Second, like Chateau and Jared (2003), syllabically-defined consistency measures (e.g., syllable 1 feedforward rime consistency) were able to account for variance in both pronunciation and lexical decision performance, with effects observed more strongly in pronunciation. It is worth noting that our results also indicate that participants are sensitive to units *other* than the sylla-

ble. Specifically, the distance consistency measure (which is *not* syllabically defined) and consistency measures that consider extrasyllabic cues are accounting for variance above and beyond that accounted for by syllabically-defined consistency.

Feedforward and feedback consistency

The correspondence between spelling and sound was operationalized via continuous measures of feedforward (spelling-to-sound) and feedback (sound-to-spelling) consistency. The effects of consistency were clearly task-modulated. For example, facilitatory effects of feedforward onset and rime consistency were present in speeded pronunciation, but not in lexical decision (see Tables 3 and 4), reflecting the pronunciation task's emphasis on converting spelling-to-sound. Interestingly, we observed feedback (but not feedforward) effects in lexical decision, which is compatible with a word recognition system where there is a bidirectional flow of activation between orthographic and phonological representations. To the extent that there is a match between phonology and orthography, there may be a type of *resonance*, allowing the system to converge more rapidly on a lexical candidate (Stone et al., 1997). The feedback consistency literature has been mixed, and reliable effects have been observed in some studies, in lexical decision (Lacruz & Folk, 2004; Perry, 2003; Stone et al., 1997) and speeded pronunciation (Balota et al., 2004; Lacruz & Folk, 2004), but these effects have not been replicated in other studies (Massaro & Jesse, 2005; Ziegler et al., 2008). Ziegler et al., in particular, have strongly argued that feedback consistency effects in the visual modality are spurious, and are most likely due to uncontrolled variables. To test for feedback consistency more stringently, we entered feedback consistency measures *after* all other surface, lexical, feedforward consistency, and semantic measures were entered. Adding feedback consistency still resulted in a small but significant increase in the variance accounted for (speeded pronunciation = 1.0%; lexical decision = 0.3%). Of course, it is quite possible that there are other variables we have not controlled for, but it is striking that feedback consistency effects remain reliable in multisyllabic word recognition, after a very large number of major covariates (more than any extant study) have been controlled for.

Given that the consistency measures accounted for a relatively modest amount of unique variance (2.4% in speeded pronunciation and 0.6% in lexical decision; see Tables 3 and 4), one could argue that consistency effects, even though significant, are not necessarily substantive. There are two responses to this. First, the amount of variance accounted for is in line with other large-scale studies. For example, in the Balota et al. (2004) monosyllabic mega-study, consistency accounted for 2.1% of variance in speeded pronunciation and 0.6% of variance in lexical decision. More crucially, the theoretical importance of consistency cannot be entirely gauged by the magnitude of its effect. For example, computational models of word recognition (Coltheart et al., 2001; Plaut et al., 1996) have been extremely influential in the literature, and yet model latencies account for very modest amounts of item-level variance in human data. In simulations reported by Spieler and Balota (1997), the Coltheart et al. model accounted for 6.9% of pronunciation variance while the Plaut et al. model accounted for only 3.1% of pronunciation variance.

Functional form of word length

New et al. (2006) were the first to report the intriguing quadratic U-shaped relationship between word length and lexical decision latencies. This is theoretically important because it suggests that there may be a preferred grain size in word recognition that is conditioned by how word length is distributed in the language. Specifically, readers appear to develop an optimal perceptual span that is about six to nine letters long, based on the most common length of words they encounter. Indeed, the mean length of words in the ELP is eight by type frequency and 7.66 by token frequency (using log HAL frequency). Words that are approximately the same length as this perceptual span appear to be processed in a more efficient (potentially parallel) manner. Words *longer* than the perceptual span are likely to engage multiple fixations; hence, the longer the word, the slower recognition should be. Finally, length may be negatively correlated with RTs for short words because very short words deviate more from the optimal span, leading to long recognition times. There is also evidence that quadratic length can be partially explained by orthographic similarity effects (Yarkoni et al., 2008). Specifically, Orthographic Levenshtein distance scores are disproportionately low for medium-length (i.e., 5–7 letters) words, because there is a preponderance of medium-length words in the language, hence shortening the orthographic distance from a target word to other words in the same length range. Indeed, Yarkoni et al. demonstrated that when Orthographic Levenshtein distance is controlled for, the unique effect of quadratic length decreases.

More importantly, the quadratic relationship indicates that reports from earlier studies examining length effects (see New et al., 2006, for a comprehensive review) are potentially misleading, due to the fact that these studies were examining words with a relatively small range of lengths. The present analyses confirm and extend New et al.'s findings by showing that the quadratic relationship is present in both pronunciation and lexical decision performance, even after a large number of correlated variables were controlled for. Moreover, the relationship was stronger in lexical decision than in pronunciation. This task effect could be due to the greater emphasis in the pronunciation task on serial left-to-right sublexical processing, which might attenuate the curvilinear trend.

Orthographic and phonological similarity

Traditionally, orthographic and phonological similarity have been investigated with orthographic and phonological N, respectively. As discussed in the Introduction, orthographic N effects have been somewhat contentious in the literature. We found facilitatory effects of this variable in both speeded pronunciation and lexical decision, and these effects were also stronger for low-frequency words than for high frequency words, consistent with Andrews (1992). Collectively, these results suggest that the presence of many orthographic neighbors help, rather than hurt,

word recognition performance, which is difficult to reconcile with the intuition that neighbors compete with each other during lexical identification (see Andrews, 1997, for a discussion of the theoretical importance of orthographic neighborhood effects).

Turning to phonological N, effects were facilitatory in speeded pronunciation, but unreliable in lexical decision. In other words, the facilitatory effects of phonological N in lexical decision performance, observed in factorial experiments using a small set of monosyllabic stimuli (Yates, 2005; Yates et al., 2004), do not generalize to performance for a large database of multisyllabic words. Interestingly, when we examined phonological N effects in monosyllabic words (see Tables 3 and 4), we observed reliable *inhibitory* effects in both speeded pronunciation and lexical decision. This unexpected finding indicates that for monosyllabic words, the presence of many phonological neighbors hurts, rather than helps, word recognition performance. What is driving the discrepancy between the present results and the work by Yates and colleagues? Possibly, not all phonological neighbors facilitate recognition, particularly when one considers shorter monosyllabic words. Interestingly, Grainger, Muneaux, Farioli, and Ziegler (2005) reported that in French lexical decision performance, phonological neighbors in dense orthographic neighborhoods were *facilitatory*, but phonological neighbors in sparse orthographic neighborhoods were *inhibitory*. Hence, it is possible that Yates' stimuli were loaded with words from dense orthographic neighborhoods, whereas most words in our large-scale analyses could have been from sparse orthographic neighborhoods. We tested this by examining the orthographic N × phonological N interaction (after controlling for onsets and lexical variables) in monosyllabic speeded pronunciation and lexical decision. Unfortunately, the interaction was not significant in speeded pronunciation, and although it was reliable in lexical decision ($p = .037$), phonological neighbors in dense orthographic neighborhoods were *more* inhibitory than phonological neighbors in sparse orthographic neighbors.

Alternatively, perhaps only words which are *both* orthographic and phonological neighbors speed up pronunciation and lexical decision performance (cf. Adelman & Brown, 2007; Peereman & Content, 1997). The other phonological neighbors may actually have no or possibly deleterious effects on performance, as demonstrated by the illustration below. For example, consider the orthographic (*cog, dig, doc, doe, don, dot, dug, fog, hog, jog, log*) and phonological (*daub, dawn, dig, doff, door, Doug, dug, hog, log*) neighbors of the word *dog*. As one can see, a number of neighbors (e.g., *hog, log*) possess similar spellings and pronunciations as the target word, *dog*. Such neighbors shorten pronunciation latencies because they activate phonology consistent with the target item's. In contrast, some phonological neighbors (e.g., *daub*, *dawn*, and *doff*) may share phonology with *dog* but relatively little overlap in orthography. In fact, some of these neighbors (e.g., *daub*) are feedback-inconsistent with the target word. That is, *dog* and *daub* share the same sound vowel, but different spelling vowels (*-o* vs. *-au*). The presence of feedback-inconsistent neighbors may actually slow down the recognition of *dog*. We explored this in our monosyllabic dataset by

replacing phonological N with *phonographic* neighborhood size in Step 3 of our regression analyses (Adelman & Brown); phonographic neighborhood size is the number of phonological neighbors that are also orthographic neighbors. When we did this, the effects of phonographic neighborhood size were facilitatory in both speeded pronunciation ($\beta = -.197$, $p < .001$) and lexical decision ($\beta = -.063$, $p = .016$). However, orthographic N effects were no longer reliable in both tasks. Interestingly, this is precisely the same pattern obtained by Ziegler and Perry (1998), who contrasted orthographic N against body N (i.e., the number of neighbors sharing the target's orthographic rime); a body neighbor might be considered a special kind of phonographic neighbor. Collectively, these results suggest that orthographic and phonological N may not be the most ideal measures of orthographic and phonological similarity, respectively.

To summarize, for monosyllabic words, it is possible that only a subset of orthographic and phonological neighbors facilitate target processing; the other neighbors either exert null or inhibitory effects. When orthographic N is entered as a predictor, its overall effect is facilitatory because there are typically more facilitating than inhibiting orthographic neighbors for a word. When phonological N alone is entered as a predictor, its overall effect is inhibitory because there are typically more inhibiting than facilitating phonological neighbors. It is possible that the Yates studies yielded facilitatory effects of phonological N because the selected stimuli featured phonological neighborhoods which were dominated by phonographic neighbors. Critically, when orthographic and phonographic neighborhood size are entered together, phonographic neighborhood size is facilitatory while orthographic N is inhibitory. This suggests that phonographic neighborhood size is accounting for all the variance explained by the facilitating orthographic neighbors, leaving the remaining orthographic neighbors to exert an inhibitory effect. Heuristically, one might think of phonographic neighborhood size as an optimized measure of orthographic *and* phonological N. To summarize, it appears that the phonological N measure could be somewhat more complex than originally conceptualized by Yates (2005). Only a subset of phonological neighbors (i.e., those which are orthographic neighbors as well) may actually be useful for helping the target word converge on its appropriate pronunciations; the other neighbors could either slow down or have no effect on target processing.

This study also explored novel measures of orthographic and phonological similarity that were based on Levenshtein distance (Yarkoni et al., 2008). Unlike standard measures of orthographic and phonological N, these distance-based measures are not only applicable to words of all lengths, but more importantly, were able to account for a substantial proportion of unique variance above and beyond the traditional measures. In general, the effects of these variables indicated that visually and phonologically confusable words were recognized faster, a counterintuitive finding that is again markedly discrepant with straightforward competitive lexical selection processes. Instead, facilitatory distance effects may reflect the processing characteristics of the sublexical phonological assembly

process, whereby words which have many words that are visually similar are recognized faster because they contain more common spelling-sound correspondences (Andrews, 1997). This argument is of course consistent with the finding that the effects of neighbors are stronger for low-frequency than for high frequency words (Andrews, 1989, 1992); low-frequency words are more likely to be influenced by sublexical processes.

Semantic effects

Semantic effects have been relatively neglected both in the empirical literature and in models of isolated word recognition. This study demonstrates that even after controlling for virtually all major lexical variables, one observes semantic effects in pronunciation and lexical decision latencies. These effects were stronger in lexical decision performance, which is consistent with the contribution of semantics to word–non-word discrimination. These results replicate Balota et al.'s (2004) study, and are again consistent with an interactive, cascadic word recognition system where semantic-level information exerts top-down influences on lower-level lexical representations relatively early in the processing stream (Balota, 1990). Of course, the semantic variables used in this study can only crudely approximate the array of semantic measures (e.g., imageability, concreteness, relatedness of meanings) that have been used to study semantic processing, and it will useful to revisit this issue when better semantic measures become available for the present corpus.

Relating effects to extant models of multisyllabic word recognition

We have described a fairly long list of targeted phenomena that a model of multisyllabic word recognition needs to accommodate. While our primary objective is to characterize *empirical* effects in multisyllabic words, it is also useful to briefly consider how well extant *multisyllabic* computational models of word recognition handle these effects. Obviously, a detailed evaluation of models is not viable, as space constraints preclude an exhaustive consideration of the underlying mechanisms and processes in each model. Nevertheless, it will be instructive to consider how well models can reproduce the item-level effects in our dataset.

Unfortunately, the answer to this question is not that straightforward, because the investigation of multisyllabic words is underrepresented in both empirical studies and model development. For example, one of the better known computational models of multisyllabic word recognition was developed by Ans, Carbonnel, and Valdois (1998) for French. The model is sensitive to statistical regularities in the language, and contains two procedures for deriving speech from print, a default global mode that operates in parallel and uses knowledge about whole words, and a secondary analytic mode that operates serially, typically syllable-by-syllable. Global processing is always engaged first, but when this fails, items will be processed by the analytic procedure. While a full description of the workings of this model is well outside the scope of this paper, it is interesting that the model contains representational units corre-

sponding to the syllable, an assumption which was motivated by the functional importance of the syllable in studies of spoken word processing. Nevertheless, however intriguing the Ans et al. (1998) model is, its architecture is based on French, which has its own idiosyncratic spelling-sound mappings. At this point, there is only one published implemented model that handles the word recognition of English multisyllabic words, the junction model of lexical processing (Kello, 2006).

Junction model of word recognition

The junction model departs significantly from DRC and connectionist frameworks by its use of a shared-route architecture (see Fig. 7), whereby orthography makes contact with phonology and semantics via the *same* representations. The outlined portion of Fig. 7 reflects the mapping between sound and meaning representations, which captures the processes and representations involved in *spoken* language acquisition and processing (i.e., converting speech to meaning). Hence, this is a single-pathway architecture, which contrasts with the DRC model, where orthography contacts phonology via either a lexical (addressed phonology) or sublexical (assembled phonology) pathway, and the connectionist model, where orthography contacts phonology either directly (Orthography → Phonology) or indirectly (Orthography → Semantics → Phonology). In general, the junction model is able to handle multisyllabic words because it can learn how to encode and decode variable-length sequences, using simple recurrent networks (SRNs) which learn the transitional probabilities of sequences. When two SRNs are joined, they form a single sequence encoder (see Fig. 8) that can convert variable-length sequences into fixed-width representations and *vice versa* (see Sibley, Kello, Plaut, & Elman, 2006, for a more detailed description). Briefly, the *input* SRN encoder is trained to convert an input of any length into a fixed-width representation, and the *output* SRN decoder is trained to recover the original input from the fixed-width representation.

This aspect of the model is critical because multisyllabic words vary greatly with respect to their number of letters and phonemes, and it is difficult to align the spelling and pronunciation of a multisyllabic word in a principled manner so as to determine their overlap (Sibley & Kello, 2006). The junction model solves the alignment problem by reducing word forms of variable length to a common
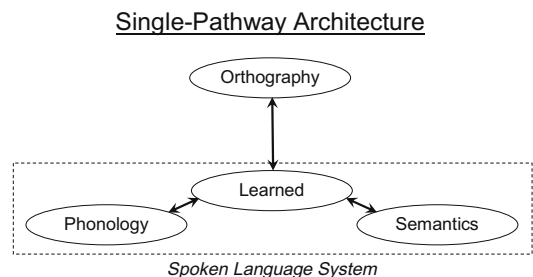

Fig. 7. The junction model of lexical processing. From Kello (2006) (p. 53). Copyright 2006 by Psychology Press Ltd. Reproduced with permission.
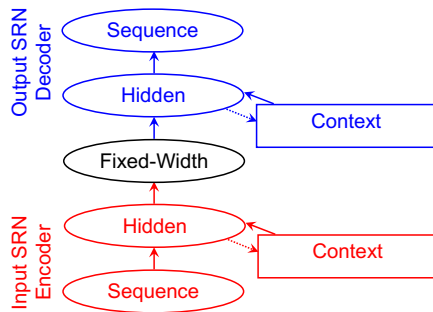
Fig. 8. Sequence encoder architecture. From Kello (2006) (p. 59). Copyright 2006 by Psychology Press Ltd. Reproduced with permission.

fixed-width base. Orthographic and phonological sequence encoders are first used to learn the orthographic and phonological representations of 45,263 words, respectively, and the textual co-occurrence statistics from the COALS database of word co-occurrences (Rohde, Gonnerman, & Plaut, 2008) were used to generate semantic codes for each word. After training, each of the 45,263 words was associated with unique 400-bit orthographic, phonological, and semantic representations. These representations were then bound together by a set of mediating representations to form the fully implemented junction model (see Fig. 9). Specifically, each of the 45,263 localist lexical nodes was bidirectionally connected to 1200 sigmoidal processing units, comprised of 400 orthographic, 400 phonological, and 400 semantic units. The functions that govern the activation dynamics of the system, and the resultant probability density function of model RTs, are outside the scope of this paper but are available in Kello (2006). It is interesting to note however, that the model produces positively skewed RT distributions, consistent with human participants.

Testing the junction model

When Kello (2006) tested the model, he reported that it could account for more variance in speeded pronunciation response times than either the dual route (Coltheart et al., 2001) or connectionist (Plaut et al., 1996) computational models. However, is the model able to reproduce the item-level effects of variables described here? For example,
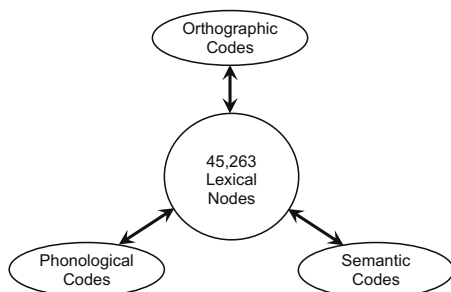

Fig. 9. The implemented junction model of lexical processing. From Kello, 2006 (p. 59). Copyright 2006 by Psychology Press Ltd. Reproduced with permission.

will rank-transformed frequency predict junction model performance better than log-transformed frequency? Fortunately, the latencies produced by the junction model are available (C. T. Kello, personal communication, December 16, 2006), making it possible to address these issues. In the next sections, two sets of analyses are described. In the first set of analyses, junction model latencies for the multisyllabic words studied in this study were regressed against log-transformed and rank-transformed frequency to assess which function accounts for more variance. Using hierarchical regression analyses, the influence of the full set of predictors used in this study on junction model latencies were then examined.

Rank or log frequency?

As discussed, recent models of lexical access are sophisticated enough to make predictions about the functional form of the frequency effect. The analyses reported by Yap (2007) indicated that the rank transformation of word frequency accounts for more variance than the more commonly used log transformation, consistent with predictions from Murray and Forster (2004, but see Adelman & Brown, 2008, for an alternative account, and Murray & Forster, 2008, for a reply). To test this, junction model latencies were correlated with log composite frequency and rank composite frequency. There were a total of 4517 monomorphemic multisyllabic words that had junction model latencies. Model RTs were more strongly correlated with rank composite frequency ($r = -.846$, $p < .001$) than with log composite frequency ($r = -.818$, $p < .001$), which is compatible with the human data. It is unclear what is responsible for the superiority of the rank transformation, since there is obviously no serial search mechanism in the junction model. Hence, the power of the rank function also appears to be quite consistent with other theoretical perspectives and is not particularly tied to a serial search mechanism, as Murray and Forster (2004) initially argued.

Item-level effects and the junction model

In this section, using the same methodology as Kello (2006), the junction model RTs were first used to predict the pronunciation and lexical decision residuals (after onsets and length were partialled out) for the monomorphemic words targeted in this study. This analysis included the 4517 monomorphemic multisyllabic words with junction model RTs and indicates how well the model can predict performance at the individual item-level (Spieler & Balota, 1997). The model accounted for 30.0% of the variance in pronunciation residuals and 38.4% of the variance in lexical decision residuals. These estimates are in fact higher than the estimates reported by Kello (2006) when he examined his full dataset of 30,894 words. Hierarchical regression analyses were conducted with junction model RTs as the dependent variable (see Table 8). The same steps used in the hierarchical analyses were included here to ensure comparability with the earlier analyses. A total of 4517 words were included in the first seven steps of the regression model. Note that 4343 words were included in the final step since not all 4517 words are represented in the two semantic measures. For comparison purposes, regres-

sion coefficients based on human data for the same items are also presented.

First, the phonological onsets accounted for virtually no variance in junction model RTs, confirming that these variables most likely reflect voice key characteristics. Second, the standard lexical variables accounted for substantially *more* variance in model RTs (72.8%) than in human performance (53.6% in speeded pronunciation and 58.8% in lexical decision). Third, the overall fit between model and human behavior was relatively good. The effects of number of syllables, word frequency, orthographic N, and the Levenshtein distance measures were reliable and in the same direction across model and human data. The model also did *not* produce significant effects of letter length, and could not correctly simulate phonological consistency or semantic effects. To summarize, the present instantiation of the junction model accounted for an impressive

proportion of pronunciation and lexical decision variance in the ELP, compared to previous attempts by computational models to accommodate monosyllabic word recognition (see Spieler & Balota, 1997; Balota & Spieler, 1998, for a discussion). Nevertheless, despite its impressive performance, it is clear that there were some interesting discrepancies between human and model performance. For example, word frequency was accounting for too much variance in junction model RTs, compared to human participants. This explains why the targeted variables could account for 73.4% of the variance in the model. Furthermore, the model was relatively insensitive to letter length and number of syllables. Of course, this is consistent with Kello's (2006) claim that the model should not be held accountable for length effects. More intriguingly, effects of phonological consistency were marginal, indicating that statistical regularities in spelling-to-sound and sound-to-

Table 8
Standardized RT and accuracy regression coefficients of the item-level regression analyses for junction model performance. Human performance (SPT = speeded pronunciation task; LDT = lexical decision task) for the same items are also presented for purposes of comparison. The *p*-value for each R^2 change is represented with asterisks.

Predictor variable	Monosyllabic words (n = 3201)			Multisyllabic words (n = 4517)			All words (n = 7718)		
	Model	SPT	LDT	Model	SPT	LDT	Model	SPT	LDT
Surface variables (onsets)									
R-square	**.015*****	**.307*****	**.004***	**.009*****	**.044*****	**.000**	**.004*****	**.066*****	**.008*****
Surface variables (stress)									
R-square	NA	NA	NA	**.025*****	**.116*****	**.051*****	**.034*****	**.179*****	**.079*****
Standard lexical variables									
Number of syllables	NA	NA	NA	.040***	.230***	.167***	.014	.258***	.203***
Length (number of letters)	.041***	.121***	−.079***	.006	.140***	.092***	.041***	.205***	.076***
Rank composite frequency	−.887***	−.386***	−.744***	−.826***	−.492***	−.660***	−.850***	−.430***	−.664***
Orthographic N	−.023*	−.204***	−.106***	−.073***	−.070***	−.054***	−.026**	−.074***	−.031*
Phonological N	−.001	.154***	.055**	−.011	−.048***	−.003	.013	.094***	.066***
R-square	**.814*****	**.526*****	**.549*****	**.728*****	**.536*****	**.588*****	**.764*****	**.586*****	**.626*****
Quadratic length									
Quadratic length	−.018	.589***	.537***	.054	.431***	.508***	.136***	.564***	.639***
R-square	**.814**	**.534*****	**.555*****	**.728**	**.540*****	**.593*****	**.765*****	**.594*****	**.637*****
Distance variables									
L orthographic distance	.023	−.039	.017	.121***	.125***	.131***	.152***	.119***	.122***
L phonological distance	.020	.230***	.086***	.047**	.267***	.192***	.066***	.342***	.200***
LOD Neighborhood frequency	.002	.122***	.122***	.028*	.130***	.099***	.037***	.162***	.127***
R-square	**.814**	**.558*****	**.563*****	**.733*****	**.584*****	**.620*****	**.771*****	**.636*****	**.657*****
Syllable 1 consistency variables									
Feedforward onset consistency	.018†	−.115***	−.025†	−.007	−.065***	−.002	.003	−.064***	−.010
Feedforward rime consistency	.012	−.060***	−.027*	.000	−.030**	.014	.014†	−.031***	−.006
Feedback onset consistency	−.013	−.070***	.009	−.003	−.017	−.007	−.004	−.019*	−.001
Feedback rime consistency	−.002	−.101***	−.006	−.002	−.059***	−.030**	−.002	−.062***	−.017*
R-square	**.814**	**.589*****	**.564†**	**.733**	**.593*****	**.621***	**.771**	**.645*****	**.657†**
Higher-order consistency variables									
Distance consistency	−.055*	−.099*	−.075†	−.089***	−.168***	−.083**	−.034*	−.055**	−.066**
Composite FF onset consistency	NA	NA	NA	.004	−.038***	−.008	.011	−.032**	−.010
Composite FF rime consistency	NA	NA	NA	.025*	−.018	−.004	.024*	−.028†	−.006
Composite FB onset consistency	NA	NA	NA	−.006	.045***	.057***	.015†	.065***	.050***
Composite FB rime consistency	NA	NA	NA	−.027*	−.094***	−.023†	−.019†	−.080***	−.015
R-square	**.815***	**.590***	**.564†**	**.734*****	**.604*****	**.624*****	**.772****	**.649*****	**.658*****
Semantic variables									
Local semantic neighborhood size	.049***	.010	.010	−.017	−.028†	−.071***	.012	−.024*	−.029**
WordNet number of senses	−.001	−.064***	−.117***	.030**	−.025*	−.059***	.023***	−.011	−.080***

* *p* < .05.
** *p* < .01.
*** *p* < .001.
† *p* < .10.

spelling correspondences are not being encoded. Finally, the model could not capture semantic effects. These could be potential areas of development in future iterations of the junction model.

The CDP++ disyllabic model

Although the junction model is technically the only *published* computational model that accommodates the reading aloud of English multisyllabic words, the monosyllabic connectionist dual-process (CDP+) model (Perry et al., 2007) has been recently extended to handle both monosyllabic and disyllabic words (Perry, Ziegler, & Zorzi, 2008), making it an exciting alternative to the junction model. The CDP+ model is an extension to the original connectionist dual process (CDP) model developed by Zorzi, Houghton, and Butterworth (1998); see Fig. 10), which has two distinct connectionist pathways that jointly support the conversion of spelling-to-sound. The monosyllabic CDP+ model (Perry et al., 2007; see Fig. 11) replaces the original retrieved phonology pathway with a fully implemented lexical pathway that is closely based on the interactive activation model (McClelland & Rumelhart, 1981), and the sublexical pathway now uses graphemes, rather than single letters, as orthographic input, a feature that has been shown to dramatically improve non-word reading accuracy (cf. Plaut et al., 1996). Finally, serial processing is hardwired into the sublexical pathway because graphemes in letter strings are individually processed in a left-to-right manner. Broadly speaking, the CDP+ model is an amalgamation of earlier computational models, effectively blending their strengths while addressing their weaknesses. For example, the DRC model (Coltheart et al., 2001) lacks the ability to learn and has difficulty capturing graded phonological consistency effects. Likewise, connectionist models (Plaut et al., 1996) are challenged by serial effects in reading (see Rastle & Coltheart, 2006, for more discussion). Simulations of the CDP+ model indicated that it could not only successfully reproduce the benchmark effects identified by Coltheart et al. (2001), but that it was also able to account for more item-level variance in databases than all earlier models. Specifically, the DRC, CDP, and connec-
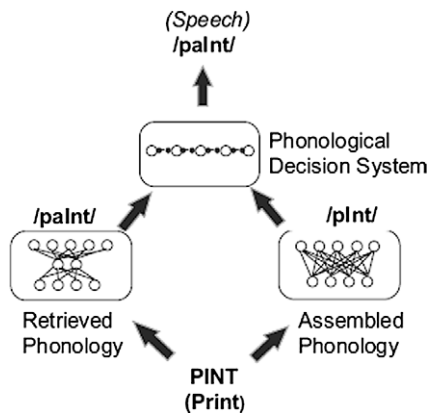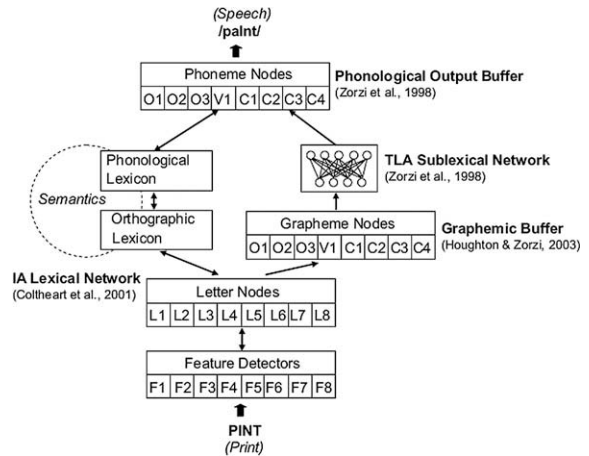

Fig. 11. The new connectionist dual process (CDP+) model. O = onset; V = vowel; C = coda; TLA = two-layer assembly; IA = interactive activation; L = letter; F = feature. From Perry et al. (2007) p. 280. Copyright 2007 by APA. Reproduced with permission.

tionist models could only account for between 3% and 7% of variance in the Spieler and Balota (1997) database, but the CDP+ model was able to account for over 17% of the variance.

It is fortuitous that Perry and colleagues have recently been working on a version of the CDP+ model that handles disyllabic words. Broadly speaking, the iteration of the disyllabic model being tested is architecturally very similar to its monosyllabic predecessor except that the grapheme buffer has been extended to code for disyllabic words (J. C. Ziegler, personal communication, December 28, 2007). Although this work is still in progress, the model's predicted latencies are available, allowing us to test the disyllabic CDP++ model against the present data set. As the model is still under development, it is important to note that some aspects, such as its parameters, may change, although we have been informed that the present instantiation of the model is quite close to the final version (J. C. Ziegler, personal communication, December 26, 2007).

Testing the disyllabic CDP++ model

The model was tested against the pronunciation and lexical decision latencies of 3394 disyllabic words for which model latencies were available. The model was able to account for 31.4% of the variance in pronunciation response times and 35.9% of the variance in lexical decision response times. Interestingly, even though the model is designed to carry out speeded pronunciation, not lexical decision, it nevertheless accounted for a remarkable amount of variance in lexical decision performance. These estimates are considerably higher than the R-squares (17.28–21.56% for speeded pronunciation performance) reported by Perry et al. (2007) when they examined monosyllabic words. CDP++ model RTs were also correlated with log and rank composite frequency, to determine which transformation of frequency had more predictive power. In contrast to the junction model and the behavioral data, log frequency (45.4%) accounted for more variance in the CDP++ model than rank frequency (42.9%).

Fig. 10. The connectionist dual process (CDP) model. From Perry et al. (2007) p. 276. Copyright 2007 by APA. Reproduced with permission.

Hierarchical regression analyses were next conducted with CDP++ model RTs as the dependent variable (see Table 9). Like the junction model, the phonological onsets accounted for virtually no variance in model RTs, which is reassuring. In addition, the target variables accounted for about the same amount of variance in model RTs (52.7%) as in human performance (54.7% in speeded pronunciation and 58.8% in lexical decision). The overall fit between model and human behavior was very good. The effects of letter length, word frequency, orthographic N, phonological N, Levenshtein phonological distance, LOD neighborhood frequency, syllable 1 feedback consistency, composite consistency, and semantic measures were reliable and in the same direction across model and human data. Number of syllables effects could not be tested with the 3394 disyllabic words but were reliably inhibitory in the full set of 6694 mono- and disyllabic words. However, unique effects of Levenshtein orthographic distance were not observed, nor did the model produce reliable effects of syllable 1 *feedforward* consistency. It is interesting that the CDP++ model correctly captured effects of number of syllables and phonological consistency. Although syllables are not explicitly represented in the model as units of representation, the grapheme buffer uses linguistic principles (such as the maximum onset principle) to determine where syllabic boundaries should lie. Disyllabic words are processed more slowly because the model initially attempts to (incorrectly) parse such words as monosyllables, and it takes time to suppress the spurious activation and to move graphemes into their correct syllabic positions (C. Perry, personal communication, December 28, 2007).

In summary, it appears that the present instantiation of the disyllabic CDP++ model is able to handle many important aspects in our dataset, with some limitations. The

Table 9
Standardized RT and accuracy regression coefficients of the item-level regression analyses for CDP++ model performance. Human performance (SPT = speeded pronunciation task; LDT = lexical decision task) for the same items are also presented for purposes of comparison. The *p*-value for each R^2 change is represented with asterisks.

Predictor variable	Monosyllabic words (n = 3300)			Disyllabic words (n = 3394)			All words (n = 6694)		
	Model	SPT	LDT	Model	SPT	LDT	Model	SPT	LDT
Surface variables (onsets)									
R-square	**.018**[***]	**.298**[***]	**.005**[**]	**.019**[***]	**.103**[***]	**.007**[***]	**.008**[***]	**.118**[***]	**.004**[***]
Surface variables (stress)									
R-square	NA	NA	NA	**.024**[***]	**.115**[***]	**.009**[**]	**.046**[***]	**.158**[***]	**.020**[***]
Standard lexical variables									
Number of syllables	NA	NA	NA	NA	NA	NA	.159[***]	.123[***]	.109[***]
Length (number of letters)	.149[***]	.120[***]	−.090[***]	.038[**]	.082[***]	.014	.113[***]	.144[***]	−.012
Rank composite frequency	−.601[***]	−.398[***]	−.740[***]	−.624[***]	−.516[***]	−.715[***]	−.563[***]	−.454[***]	−.715[***]
Orthographic N	−.017	−.202[***]	−.110[***]	−.030[†]	−.114[***]	−.092[***]	−.031[*]	−.133[***]	−.075[***]
Phonological N	−.041[*]	.147[***]	.049[**]	−.179[***]	−.060[***]	−.007	−.091[***]	.073[***]	.042[**]
R-square	**.451**[***]	**.529**[***]	**.546**[***]	**.479**[***]	**.440**[***]	**.537**[***]	**.571**[***]	**.513**[***]	**.581**[***]
Quadratic length									
Quadratic length	−.052	.642[***]	.541[***]	.521[***]	.871[***]	.671[***]	.067	.685[***]	.683[***]
R-square	**.451**	**.537**[***]	**.553**[***]	**.483**[***]	**.451**[***]	**.544**[***]	**.571**	**.524**[***]	**.591**[***]
Distance variables									
L orthographic distance	.004	−.006	.013	−.006	.085[***]	.113[***]	−.014	.077[***]	.101[***]
L phonological distance	.226[***]	.207[***]	.093[***]	.246[***]	.283[***]	.173[***]	.317[***]	.367[***]	.187[***]
LOD neighborhood frequency	.021	.121[***]	.119[***]	.071[***]	.159[***]	.109[***]	.068[***]	.181[***]	.136[***]
R-square	**.470**[***]	**.559**[***]	**.561**[***]	**.518**[***]	**.525**[***]	**.581**[***]	**.600**[***]	**.583**[***]	**.615**[***]
Syllable 1 consistency variables									
Feedforward onset consistency	−.027[†]	−.093[***]	−.031[*]	−.011	−.070[***]	−.023[†]	−.018[†]	−.067[***]	−.025[**]
Feedforward rime consistency	−.120[***]	−.058[***]	−.022[†]	−.004	−.036[**]	.017	−.060[***]	−.057[***]	−.003
Feedback onset consistency	−.005	−.086[***]	.000	−.050[**]	−.036[*]	−.006	−.024[*]	−.041[***]	−.002
Feedback rime consistency	−.001	−.097[***]	-.014	−.034[**]	−.050[***]	−.014	−.025[**]	−.058[***]	−.009
R-square	**.482**[***]	**.588**[***]	**.562**[*]	**.521**[***]	**.535**[***]	**.581**	**.604**[***]	**.596**[***]	**.615**[*]
Higher-order consistency variables									
Distance consistency	−.190[***]	−.133[***]	−.056	−.040	−.167[***]	−.108[**]	−.131[***]	−.070[**]	−.060[*]
Composite FF onset consistency	NA	NA	NA	−.075[***]	−.032[*]	.003	−.065[***]	−.022[†]	.007
Composite FF rime consistency	NA	NA	NA	−.053[**]	−.050[**]	−.012	−.099[***]	−.043[*]	−.010
Composite FB onset consistency	NA	NA	NA	.057[***]	.060[***]	.042[**]	.084[***]	.065[***]	.038[**]
Composite FB rime consistency	NA	NA	NA	−.042[*]	−.113[***]	−.045[**]	−.025	−.116[***]	−.048[**]
R-square	**.485**[***]	**.589**[***]	**.562**	**.527**[***]	**.547**[***]	**.584**[***]	**.610**[***]	**.601**[***]	**.616**[***]
Semantic variables									
Local semantic neighborhood size	.041[*]	.007	.018	−.321[***]	−.037[†]	−.086[***]	−.081[***]	−.028[**]	−.026[*]
WordNet number of senses	−.127[***]	−.066[***]	−.121[***]	−.101[***]	−.047[***]	−.060[***]	−.113[***]	−.035[***]	−.089[***]

[*] *p* < .05.
[**] *p* < .01.
[***] *p* < .001.
[†] *p* < .10.

model mimicked many behavioral effects correctly, but it did not show the rank frequency advantage, nor did it produce effects of Levenshtein orthographic distance and syllable 1 feedforward consistency. The significant feedback consistency and semantic effects shown by the model are intriguing and puzzling, given that the CDP++ model does not implement sublexical feedback from phonology to orthography and lacks semantic representations. It is possible that these model effects are spuriously driven by uncontrolled covariates, or from suppressor and interactive effects in a set of highly intercorrelated variables. Whether these factors are similarly responsible for the human effects is a question that merits further investigation. More importantly, the assumption that *every* significant effect necessarily implicates a corresponding representation and process may be too simplistic, and should be tempered by the abovementioned considerations.

Limitations of present study and future directions

Although this study has examined the influence of a large number of measures on visual word recognition performance, a number of questions remain unanswered. For example, although large, the number of orthographic, phonological, and semantic variables is obviously not exhaustive. Specifically, morphologically-defined measures (e.g., part of speech, number of complex words in which a target word occurs as a constituent) have also been shown to play a prominent role in visual word recognition (Baayen et al., 2006), and it is important to understand how morphological effects operate in multisyllabic word recognition. A better understanding of stress assignment processes also remains an important target for future research.

Second, the analyses described here generally emphasize the *linear* effects of variables. However, psycholinguistic variables are not always linearly related to RTs. For example, both log frequency and length produced curvilinear effects. It is important to identify other variables that may be producing such non-linear relationships with RTs, and how these non-linear relationships interact with other variables.

Third, since the primary objective of this study was to identify benchmarks that could be used to constrain multisyllabic word recognition models, the *empirical* aspects of the data have been emphasized. Of course, the theoretical ramifications were also discussed to a lesser or greater extent throughout, but given the theoretical development of various models regarding the targeted multisyllabic variables, further speculation is premature at this point. However, some questions clearly merit more investigation. For example, why does rank-transformed frequency have so much predictive power, and what implications does this have for models of lexical access and more generally learning? How can input coding mechanisms be constrained by the surprisingly robust Levenshtein distance effects, where words which are more orthographically and phonologically confusable are easier to recognize? Why does the CDP++ model produce semantic and feedback phonology effects, even though the mechanisms hypothesized to underlie these effects are not implemented?

Finally, although this paper has been represented as a *multisyllabic* word recognition megastudy, the reality is that monomorphemic words form only a modest quarter of the 40,481 words in the ELP. The extent to which the effects reported in this study generalize to these multimorphemic mono- and multisyllabic words remains an open empirical question, and is clearly an important next step for the field.

Concluding remarks

The present study examined the influence of various variables on the pronunciation and lexical decision latencies of 6115 monomorphemic multisyllabic words. The results of the regression analyses indicated that the constellation of parameters accounted for a remarkable 61.2% and 61.6% of the variance in pronunciation and lexical decision performance, respectively, which is clearly considerably higher than the estimates reported in previous studies. The present study has broken ground on two additional fronts. First, it is currently the largest megastudy of multisyllabic word recognition utilizing a comprehensive array of traditional and novel orthographic, phonological, and semantic predictors. There are other multisyllabic word recognition megastudies in the literature but they either studied fewer words (Chateau & Jared, 2003) or controlled for a relatively small number of lexical variables (New et al., 2006). Second, the item-level regression analyses accounted for over 60% of the variance in both pronunciation and lexical decision latencies, substantially more than any previous word recognition megastudy. Hopefully, the results reported here will provide useful constraints for emerging models of multisyllabic visual word recognition. This is timely as researchers (e.g., Kello, 2006; Perry et al., 2008) have begun to develop multisyllabic word recognition models. It is important that these models capture not only the item-level variance associated with human pronunciation and lexical decision latencies, but also reproduce the constellation of effects shown by human participants. Our initial examination of such models indicate impressive power, with, as expected, some limitations that should serve as targets in future model development.

References

Adelman, J. S., & Brown, G. D. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*, 455–459.

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115*, 214–227.

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

Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*, 439–461.

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multi-trace memory model of polysyllabic word reading. *Psychological Review, 105*, 678–723.

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*, 920–948.

Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 34*, 224–236.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory & Language, 55*, 290–313.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A. (1990). The role of meaning in word processing. In D. A. Balota, G. Flores D'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9–32). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance, 10*, 340–357.

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*, 283–316.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition, 29*, 639–647.

Balota, D. A., & Spieler, D. H. (1998). The utility of item-level analyses in model evaluation: A reply to Seidenberg and Plaut. *Psychological Science, 9*, 238–240.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 285–375). Amsterdam: Academic Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., et al (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–244). Basil Blackwell: London.

Brand, M., Rey, A., & Peereman, R. (2003). Where is the syllable priming effect in visual word recognition? *Journal of Memory & Language, 48*, 435–443.

Brants, T., & Franz, A. (2006). *Web IT 5-gram version 1*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition, 7*, 68–76.

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

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7*, 249–253.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates, Publishers.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.). *Attention and performance* (VI). Hillsdale, NJ: LEA.

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*, 204–256.

Content, A., Kearns, R., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory & Language, 45*, 177–199.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments & Computers, 36*, 384–387.

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

Cutler, A. (1981). Making up materials is a confounded nuisance: Or will we be able to run any psycholinguistic experiments in 1990? *Cognition, 10*, 65–70.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory & Language, 25*, 385–400.

Durda, K., Buchanan, L., & Caron, R. (2006). WordMine2. University of Windsor. <http://www.wordmine2.org>.

Ferrand, L. (2000). Reading aloud polysyllabic words and nonwords: The syllabic-length effect re-examined. *Psychonomic Bulletin & Review, 7*, 142–148.

Ferrand, L., & New, B. (2003). Syllabic length effects in visual word recognition and naming. *Acta Psychologica, 113*, 167–183.

Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. *Memory & Cognition, 25*, 458–470.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity, and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*, 256–281.

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

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighborhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 58A*, 981–998.

Hooper, J. B. (1972). The syllable in phonological theory. *Language, 48*, 525–540.

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

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

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

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human Perception & Performance, 16*, 92–105.

Kello, C. T. (2006). Considering the junction model of lexical processing. In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 50–75). New York: Psychology Press.

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

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

Kessler, B., Treiman, R., & Mullennix, J. (2008). Feedback consistency effects in single-word reading. In E. J. Grigorenko & A. Naples (Eds.), *Single-word reading: Biological and behavioral perspectives* (pp. 159–174). Mahwah, NJ: Erlbaum.

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, 57A*, 1261–1284.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28*, 203–208.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Massaro, D. W., & Jesse, A. (2005). The magic of reading: Too many influences for quick and easy explanations. In T. Trabasso, J. Sabatini, D. W. Massaro, & R. C. Calfee (Eds.), *From orthography to pedagogy: Essays in honor of Richard L. Venezky* (pp. 37–61). Mahwah, NJ: Erlbaum.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–407.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography, 3*, 235–312.

Monaghan, J., & Ellis, A. W. (2002). What, exactly, interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology: Learning Memory & Cognition, 28*, 183–206.

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 799–810.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111*, 721–756.

Murray, W. S., & Forster, K. I. (2008). The rank hypothesis and lexical decision: A reply to Adelman and Brown (2008). *Psychological Review, 115*, 240–252.

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review, 13*, 45–52.

Patterson, K. E., & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia* (pp. 335–359). Hillsdale, NJ: Erlbaum.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory & Language, 37*, 382–410.

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 & Language, 39*, 151–174.

Perry, C. (2003). A phoneme–grapheme feedback consistency effect. *Psychonomic Bulletin & Review, 10*, 392–397.

Perry, C., Ziegler, J. C., & Zorzi, M. (2008). Beyond single syllables: Simulating reading aloud of disyllabic words with CDP+. In *Forty-ninth annual meeting of the psychonomic society*, Chicago, IL [November].

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*, 273–315.

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*, 56–115.

Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory & Language, 25*, 461–475.

Rapp, B. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory & Language, 31*, 33–53.

Rastle, K., & Coltheart, M. (2006). Is there serial processing in the reading system; and are there local representations? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing* (pp. 3–24). Hove, UK: Psychology Press.

Rohde, D. L. T., Gonnerman, L. M., & Plaut, D. C. (2008). *An improved method for deriving word meaning from lexical co-occurrence*. Unpublished manuscript. Available from dlt4.mit.edu/~dr/COALS/Coals.pdf.

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 512–528.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 876–900.

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart's (Ed.). *Attention and performance: The psychology of reading* (Vol. XII, pp. 245–263). Hillsdale, NJ: Erlbaum.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

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*, 383–404.

Selkirk, E. O. (1982). The syllable. In H. Van der Hulst & N. Smith (Eds.), *The structure of phonological representations (Part II)*. Dordrecht: Foris.

Sibley, D. E., & Kello, C. T. (2006). Learning representations of orthographic wordforms. In *Proceedings of the 27th annual meeting of the cognitive science society*. Vancouver, BC, Canada.

Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2006). Large-scale modeling of wordform learning and representation. *Cognitive Science, 32*, 741–754.

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

Spoehr, K. T. (1981). Word recognition in speech and reading: Toward a single theory of language processing. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum.

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 & Language, 36*, 337–359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1140–1154.

Taft, M. (1979). Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior, 18*, 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 & Performance, 18*, 1004–1014.

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*, 107–136.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50A*, 439–456.

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, MO.

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979.

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1385–1397.

Yates, M., Locker, L., & Simpson, G. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review, 11*, 452–457.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science.

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.

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

Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition, 68*, 53–62.

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*, 643–661.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 1131–1161.