Running head: AUDITORY WORD RECOGNITION

Keywords: auditory word recognition; spoken word recognition; auditory lexical decision; large-scale studies; megastudies; word frequency; age of acquisition; neighborhood density; onset duration

Auditory word recognition of monosyllabic words: Assessing the weights of different factors in lexical decision performance

Melvin J. Yap[1] & Marc Brysbaert[2]

Department of Psychology, National University of Singapore[1]

Ghent University, Belgium[2]

Address correspondence to: Melvin J. Yap
Department of Psychology, Faculty of Arts and Social Sciences
National University of Singapore
Block AS4, #02-07
Singapore 117570, Republic of Singapore
melvin@nus.edu.sg

Abstract

The literature on auditory word recognition has been dominated by experimental studies, where researchers examine the effects of dichotomized variables (e.g., frequency) on response times and accuracy, while controlling for extraneous variables. Although experiments help inform and constrain the lexical processing system in important ways, they are also associated with some limitations. In the present study, we explore the utility of analyzing existing datasets via regression analyses, in order to complement and extend findings from experimental work. Specifically, using three independent auditory lexical decision datasets, we evaluated the relative importance of onset characteristics, token duration, word frequency, neighborhood density, uniqueness point, consistency, imageability, and age of acquisition (AoA) on response times and accuracy. Surprisingly, onset characteristics, duration, and AoA accounted for more item-level variance than predictors that are far more influential in the literature. The discussion focuses on the new theoretical and methodological insights provided by these analyses.

Auditory word recognition of monosyllabic words: Assessing the weights of different factors in lexical decision performance

Language processing is critical to human life. Therefore, it is no coincidence that language research plays a central role in cognitive psychology. Major efforts have been made to understand how people produce and perceive spoken and written messages. In this article we present a review of the factors claimed to influence the efficiency with which spoken words are recognized. In particular, we review the literature on the auditory lexical decision task. In this task, participants are presented with spoken stimuli and they have to decide whether the stimuli form a word or not. The lexical decision task is one of the most popular tasks to study word processing, both in the auditory and the visual modality. The availability of data in both modalities makes it possible not only to assess the impact of the different variables on spoken lexical decision, but also to compare them with what happens in visual lexical decision. First, we review the various factors that have been mentioned to influence auditory lexical decision performance. Then, we report new analyses that will allow us to assess their relative weights.

Variables influencing auditory lexical decision performance

In a systematic and thoughtful review of the auditory lexical decision task, Goldinger (1996) highlighted several variables that were known to affect lexical decision performance[1]. The variable consistently reported to have an effect was *word frequency*; high-frequency words are recognized faster than low-frequency words. This is unsurprising given the ubiquity of the frequency effect, both in visual word recognition (Baayen, Feldman, & Schreuder, 2006; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Brysbaert & New, in press) and in auditory word recognition (for lexical decision, see Cleland, Gaskell, Quinlan, & Tammimen, 2006;

Connine, Mullenix, Shernoff, & Yelen, 1990; Goh, Suárez, Yap, & Tan, in press; Luce & Pisoni, 1998; Marslen-Wilson, 1990; Meunier & Segui, 1999; Slowiaczek & Pisoni, 1986; Taft & Hambly, 1986; for phoneme categorization, see Connine, Titone, & Wang, 1993; and for word naming, see Luce & Pisoni, 1998). According to different models of spoken word recognition, word frequency can modulate the recognition threshold (e.g., the logogen model; Morton, 1969) or the resting activation of lexical representations (e.g., the cohort model; Marslen-Wilson, 1987). It can also influence the strength of connections between lexical and sublexical representations (MacKay, 1982, 1987) and bias postlexical decision processes (neighborhood activation model; Luce & Pisoni, 1998). Word frequency appears to affect performance even at the earliest moments of recognizing a spoken word (Cleland et al., 2006). Specifically, when Cleland et al. used a dual-task procedure (Pashler, 1994) to identify the locus of the frequency effect in auditory word recognition, they concluded that frequency-sensitive processes in auditory word recognition are not only automatic but also operate early (but see Connine et al., 1993, for a contrasting view).

The second variable Goldinger (1996) mentioned was *neighborhood density* (e.g., Luce & Pisoni, 1998, Vitevitch, 2002): Words which are more phonologically distinct are recognized faster than words that are less distinct. This is due to the fact that auditory stimuli activate partially compatible word representations in addition to the target words, so that there is competition between the various representations that get activated. This competition is most often captured with neighborhood density, which refers to the number of neighbors a target word has. There are two metrics for density. Density A counts only neighbors that are defined as words that can be obtained by substituting one phoneme of the word (e.g. the neighbors of *hatch* include *<u>m</u>atch*, *h<u>i</u>tch*, and *ha<u>v</u>e*). Density B in addition includes neighbors obtained by deleting one phoneme (e.g., *as* for *has*) or by adding one phoneme (e.g., *halves* for *has*). The latter definition[2] seems to be more influential; Luce and Pisoni (1998) used it and it is becoming more frequent in visual word recognition as well (De Moor & Brysbaert, 2000; Davis & Taft, 2005). At the time of Goldinger's review, there was only one study showing an effect of neighborhood density (Luce, 1986) and one that failed to obtain it (Marslen-Wilson, 1990). Since the review, however, a series of studies have confirmed the importance of neighborhood density. Luce and Pisoni (1998), for instance, reported that word frequency alone explained at most 6% of the variance in auditory perceptual identification tasks, whereas a frequency-weighted neighborhood probability measure, taking into account both the frequency of the target and the frequencies of the competitors, explained up to 22% of the variance. These trends were also observed in the auditory lexical decision task.

Third, Goldinger (1996) mentioned the *uniqueness point*, which refers to the position in a word that distinguishes the word from all other words. Given that spoken words take time to produce, they can often be recognized before they are fully pronounced, certainly when the first part uniquely defines the word (as in *spaghetti*). As a consequence, it can be expected that words with earlier uniqueness points will be recognized faster than words with later uniqueness points, everything else being equal. The uniqueness point is an important variable within the cohort theory (Marslen-Wilson, 1989; Marslen-Wilson & Welsh, 1978), which is based on the assumption that the auditory input initially activates all words compatible with the first segment and subsequently prunes back the number of candidates until only the target word remains. Although the uniqueness point can be defined in two ways (i.e., as the number of phonemes up to the uniqueness point or the time duration up to this point), it is most typically defined in terms of the number of phonemes.

Finally, Goldinger (1996) mentioned the importance of matching stimuli on *stimulus length*. Given that some words take longer to pronounce than others, it seems wise to make sure that words in one condition are not systematically longer than those in the other condition(s). Again, there are two ways to define stimulus length, either as the number of phonemes in the word, or as the actual duration of the word. The former is usually mentioned in articles. The latter is often controlled implicitly when researchers start the response timer from stimulus *offset* (instead of onset) or when they match the stimuli across the various conditions on token duration.

A variable that has gained prominence after Goldinger's (1996) review is the *consistency* of the mapping between the pronunciation and the spelling of the words, which reflects the extent to which words with similar pronunciations have similar spellings (Stone, Vanhoy, Van Orden, 1997). Consistent words (e.g., *cad*) have spellings that match that of similarly pronounced words (e.g., *bad*, *dad*, *pad*), while inconsistent words (e.g., *scheme*) have spellings that are in conflict with similar pronounced words (e.g., *dream*). While consistency can be computed for various orthographic segments, it has most often been examined at the level of the rime (i.e., medial vowel plus coda). There is substantial evidence in the literature that consistent words are recognized faster than inconsistent words in auditory word recognition, suggesting that orthographic information modulates the perception of speech, although most of this research has been done in French (Pattamadilok, Morais, Ventura, & Kolinsky, 2007; Perre & Ziegler, 2008; Ziegler & Ferrand, 1998; Ziegler, Ferrand, & Montant, 2004; Ziegler & Muneaux, 2007; Ziegler et al., 2008) or Portuguese (Ventura, Morais, & Kolinsky, 2007; Ventura, Morais, Pattamadilok, & Kolinsky, 2004). Most English studies (e.g., Chereau, Gaskell, & Dumay, 2007; Miller & Swick, 2003; Slowiaczek, Soltano, Wieting, & Bishop, 2003; Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008) have examined primed rather than isolated word recognition by manipulating the orthographic and phonological overlap between prime and target. To our knowledge, the only study that reported consistency effects in English auditory lexical decision, without using priming, is Ziegler, Petrova, and Ferrand (2008).

Another variable that has been added to the series of possible influences on auditory word recognition is the *age of acquisition* (AoA; i.e., the age at which a word has been learned). For example, Turner, Valentine, and Ellis (1998) showed that words acquired early in life are recognized faster than words acquired later, independent of their frequency of occurrence. Intriguingly, these authors even claimed that that when AoA was controlled for, frequency no longer had a reliable effect on auditory lexical decisions, suggesting that previous studies reporting robust frequency effects may have inadvertently reported an AoA effect in disguise (see Smith, Turner, Brown, & Henry, 2006, for a similar finding of a stronger AoA effect than a frequency effect).

Other researchers have explored the role of meaning-related variables on the perception of speech. For example, both Tyler, Voice, and Moss (2000) and Wurm, Vakoch, and Seaman (2004) demonstrated that high-*imageability* words (i.e., words that are easy to visualize) are recognized faster than low-imageability words. Wurm et al. (2004) also showed effects of other meaning-related variables such as Osgood’s (1969) dimensions of Evaluation (“is this good or bad?”), Activity (“is it fast or slow?”), and Potency (“is it strong or weak?”).

The value of multiple regression analyses

When confronted with a considerable list of variables like the one just mentioned, it is a good strategy to both consider the effect of each variable in isolation and to run multiple regression analyses on (large) samples of unselected stimuli. This approach has been applied successfully in visual word recognition and is known there as the megastudy approach (see

Baayen et al., 2006; Balota et al., 2004; Chateau & Jared, 2003; Cortese & Khanna, 2007; Lemhofer, Dijkstra, Schriefers, Baayen, Grainger, & Zwitserlood, 2008; Lewis & Vladeanu, 2006; Seidenberg & Waters, 1989; Spieler & Balota, 1997; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995; Yap & Balota, 2009). Although these regression studies are probably not as powerful as carefully designed experiments to evaluate the significance of specific variables or combination of variables (Sibley, Kello, & Seidenberg, 2009), they provide invaluable information about the relative importance of multiple correlated variables in terms of the percentage of word recognition variance explained.

Another reason for running regression analyses on unselected stimulus samples rather than trying to manipulate particular variables was mentioned by Lewis and Vladeanu (2006). They pointed to the fact that in lexical processing research, experimenters are rarely able to manipulate their variables. All psycholinguists can do is select the stimuli for the different conditions. As Lewis and Vladeanu (2006, p. 979) argued: "… we cannot manipulate the factors that we call independent variables. A word has high frequency, not because we have manipulated its frequency, but because there is something that causes it to be produced more often than other words. When we identify word frequency effects, these effects are merely correlations between two dependent measures." The fact that word stimuli have to be selected (rather than manipulated) strains the solidity of experimental results because researchers usually have difficulty matching their stimuli on all control variables identified in the literature (Cutler, 1981) and because there is always the danger of subtle biases in the type of words that are chosen for the different conditions (Forster, 2000).

Finally, the fact that experimenters categorize continuous variables in experiments is another restriction, given that such categorization is associated with a decrease in statistical power (Cohen, 1983; Humphreys, 1978; Maxwell & Delaney, 1993) or sometimes an increased likelihood of incorrectly rejecting the null hypothesis (MacCallum, Zhang, Preacher, and Rucker, 2002).

The following are some of the insights yielded by regression analyses on visual lexical processing data:

1. Word frequency is the most important predictor of visual lexical decision times, accounting for up to 40% of the variance (of which 25% cannot be accounted for by other correlated variables; Baayen et al., 2006; Cortese & Khanna, 2007). In contrast, for word naming times, the articulatory features of the initial phoneme are the most important, explaining up to 35% of the variance (Balota et al., 2004; Cortese & Khanna, 2007). In this task, word frequency explains less than 10% of the variance (of which 6% is pure), implying that for word naming it is more critical to match conditions on the first phoneme than on frequency (Kessler, Treiman, & Mullennix, 2002; Rastle, Croot, Harrington, & Coltheart, 2005; Rastle & Davis, 2002).
2. There are large quality differences between various word frequency measures. In particular, the widely used Kučera and Francis (1967; KF67) frequency norms are bad. The proportion of variance explained by KF67 frequency in visual lexical decision times is more than 10% less than the variance explained by the best available frequency estimates (Balota et al., 2004; Brysbaert & New, in press; Zevin & Seidenberg, 2002).
3. When objective frequency, familiarity ratings, and age-of-acquisition (AoA) ratings are used as predictors, the total proportion of variance explained for monosyllabic printed words remains the same regardless of the quality of the objective frequency measure used

(10% in word naming and 51% in lexical decision times respectively; Brysbaert & Cortese, submitted). This is due to a trade-off between the objective frequency measure and the familiarity measure (and to a lesser extent the AoA rating). Specifically, the *more* variance accounted for by objective frequency, the *less* variance accounted for by familiarity. Rated familiarity and rated AoA account for 20% of the visual lexical decision times when KF is used, but for less than 5% if the best objective frequency measure is used.

4. There is a quadratic effect of word length in visual lexical decision if word frequency and neighborhood density are controlled for: RTs decrease for very short word lengths (2-4 letter), stay stable for middle word lengths (5-8 letters), and increase sharply after that (9+ letters; New, Ferrand, Pallier, & Brysbaert, 2006).
5. Many theoretically important variables account for at most 3% of the variance in lexical decision times to monosyllabic printed words, of which usually less than 1% is unquestionably due to these variables (Baayen et al., 2006).
6. The best predictor of word processing times of English words in a second language is the processing times of English words in the native language. The specific properties of the mother tongue of bilinguals account for very little variance, putting into perspective the many experimental studies that focus on the interactions between the bilinguals' first and second languages (Lemhofer et al., 2008).

The examples selected above illustrate a few of the contributions regression analyses have made to our understanding of *visual* word processing. In the remainder of this article, we will examine how regression analyses can yield additional insights into *auditory* word processing (see Jusczyk and Luce, 2002, and Dahan & Magnuson, 2006, for excellent reviews of the literature). Our emphasis will be on performance in the auditory lexical decision task. While other auditory lexical processing tasks are available, such as word naming, where participants repeat auditorily presented words, or perceptual identification, where participants identify words that are degraded by presenting them against a background of white noise, the lexical decision task has been particularly influential in the literature for several reasons. It is easy to administer, produces more robust effects than naming, does not require the use of degraded stimuli, is less susceptible to sophisticated guessing strategies, and allows response times to be measured (Luce & Pisoni, 1998). Much of the research using the auditory lexical decision task has examined the effects of cohorts (Taft & Hambly, 1986; Marslen-Wilson, 1990; Soares, Collet, & Duclaux, 1991) or competing neighbors (Goh et al., in press; Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998; Vitevitch & Luce, 1999; Ziegler, Muneaux, & Grainger, 2003), the effects of various types of priming (e.g., Holcomb & Anderson, 1993; Kiyonaga, Grainger, Midgley, & Holcomb, 2007; Whatmough, Arguin, & Bub, 1999), and the involvement of orthographic representations in spoken word recognition (Ziegler & Ferrand, 1998; Ziegler et al., 2003; Ziegler et al., 2008).

In particular, we will be focusing on the relative weights of the variables mentioned above and on how they compare to what is found in visual lexical decision. Multiple regression analyses have been used before to study auditory lexical decision performance (Smith et al., 2006; Wurm et al., 2004), but they have focused on the *significance* of the variables, not on their *importance* vis-à-vis each other.

Method

Datasets

To conduct regression analyses, one needs reasonably large samples of items. In visual word recognition, studies based on large sets of items have become known as megastudies, where stimulus sample sizes range from a few hundred to over 40,000 words. Interestingly, with the exception of Luce and Pisoni's (1998) seminal study, megastudies are virtually non-existent in auditory word recognition research, very likely because presenting auditory stimuli entails a great deal more effort than presenting visual stimuli. However, in the literature it is possible to find studies involving a few hundred stimuli and our previous experiences with regression-type analyses have convinced us that sensible conclusions can be drawn from these samples, unless the range of the stimuli is restricted in some way. In addition, most auditory lexical decision studies thus far have involved monosyllabic words, which means that the maximum number of stimuli is limited to some 8,000 (even less if only monomorphemic words are considered).

We were able to obtain item-level data from three studies[3]. The first is the well-known large-scale study of Luce and Pisoni (1998, Experiment 2), in which they tested their Neighborhood Activation model (Luce, 1986; Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990). This study involved 918 consonant-vowel-consonant (CVC) monosyllabic words that differed on KF67 frequency, neighborhood density, and neighborhood frequency. The second study consisted of an unpublished pilot study referred to in Cleland et al. (2006). It consisted of 200 low-frequency and 199 high-frequency monosyllabic words (frequencies based on the CELEX database; Baayen, Piepenbrock, & van Rijn, 1993). Finally, we had access to a small-scale study run by Goh et al. (in press). This study, which factorially manipulated neighborhood density and word frequency, consisted of 184 monosyllabic CVC words. Because the vast majority of stimuli (88%) consisted of three phonemes, there was no point in entering phoneme length in the regression analyses.

Predictor variables

Word frequency. For a long time, the KF67 frequency measure, based on a small corpus of one million words, was the only index available for English. Even now, it is often used to investigate the impact of word frequency, even though other, better measures have become available. The first real alternative was CELEX frequency (Baayen et al., 1993), which is based on 16.6 million written and 1.3 million spoken words. Next came Zeno frequency (Zeno, Ivens, Millard, & Duvvuri, 1995), based on 17 million words from school books going from grade 1 to grade 12. Another interesting addition was the British National Corpus frequency (BNC, Leech, Rayson, & Wilson, 2001), containing 89.7 million words from written sources and 10.6 million words from spoken sources.

The advent of the Internet has spurred the development of frequency counts based on millions of words downloaded from websites and discussion groups. The first of these was the Hyperspace Analog to Language frequency (HAL, Conley, Burgess, & Hage, 1999; Lund & Burgess, 1996), based on a corpus of more than 300 million words. The second, USENET frequency, has been regularly updated and the most recent iteration is based on 11 billion words (Shaoul & Westbury, 2009). Probably the most impressive Internet frequency counts to date are the Google frequency norms, derived from approximately one trillion words[4] gathered from publicly accessible webpages (Brants & Franz, 2006). Specifically with respect to spoken words, in addition to CELEX and BNC, there are two small corpora based on auditory language. The first contains 1.6 million words coming from transcriptions of lectures, meetings, advisement sessions, and public addresses at a university (Pastizzo and Carbone, 2007). The second is the spoken American National Corpus frequency (ANC_{spoken}, Ide & Macleod, 2001: 3.86 million). The small size of the spoken corpora is due to the costs associated with speech transcription.

The latest addition to frequency measures comes from Brysbaert and New (in press). They presented a $SUBTL_{WF}$ frequency measure, which estimates word use on the basis of film and television subtitles (51 million words). In addition to raw frequency counts, this database also provides a measure of contextual diversity, $SUBTL_{CD}$. This is a measure that indicates in how many films a particular word is used (rather than the total number of observations of the word). Adelman, Brown, and Quesada (2006) have argued that contextual diversity measures better explains word processing efficiency, a finding replicated by Brysbaert and New (in press) for several megastudies involving visual word stimuli.

In the Results section, we will examine the correlations between the various frequency measures and the auditory lexical decision performance indices (reaction times and percentage of errors) and select the best frequency measure on the basis of this analysis. In particular, we report data on the following measures (we also tested the others, but they did not alter the conclusions we draw):

1. KF67 (obtained from http://elexicon.wustl.edu/; verified on September 8, 2009)[5]
2. CELEX (available at http://celex.mpi.nl/; verified on September 8, 2009). There are measures both from written and spoken sources.
3. Zeno (Zeno et al., 1995). This measure is not freely available, but is included here because it has repeatedly been shown to be one of the best for visual word recognition.
4. HAL (available at http://elexicon.wustl.edu/; verified on September 8, 2009)

5. British National Corpus (available at http://www.kilgarriff.co.uk/bnc-readme.html; verified on September 8, 2009). Only the spoken part will be tested.
6. $SUBTL_{WF}$ and $SUBTL_{CD}$ (available at http://expsy.ugent.be/subtlexus; verified on September 8, 2009). The WF measure counts the number of occurrences in the corpus of 51 million words; the CD measure counts the number of films in which the word is used (out of a total of 8,388).

Neighborhood density. We downloaded the densities as defined by Luce and Pisoni (1998) from the 20,000-word Hoosier Mental Lexicon database (obtained from http://128.252.27.56/Neighborhood/Home.asp, retrieved on August 10, 2009). As mentioned in the Introduction, there are two metrics for density. Density A counts only neighbors that are defined as words that can be obtained by substituting one phoneme of the word, whereas density B in addition includes neighbors obtained by deleting one phoneme. Because both metrics are highly correlated (r = .97) and because density B is generally seen as more appropriate, our analyses were based on this measure.

Uniqueness point. As discussed, uniqueness point is measured from the beginning of a word and refers to the point (i.e., position of phoneme) where a word diverges from all other words in the lexicon. We used the estimates computed in Luce (1986), which were provided by the Luce group (M. Geer, personal communication, September 3, 2009).

Stimulus duration. Stimulus duration refers to the duration of the recorded token in ms for each word.

Consistency. Consistency specifically refers to token (i.e., weighted by frequency) feedback rime consistency. For example, the token feedback rime consistency of *half* is computed by dividing the summed log frequencies of friends (i.e., words that have rimes pronounced as /□f/ and spelled as –alf, e.g., *calf*) by the summed log frequencies of friends *and* enemies (i.e., words that have rimes pronounced as /□f/ but not spelled as –alf, e.g., *graph*). Consistency values range from 0 (least consistent) to 1 (most consistent), and our analyses are based on the measure developed by Balota et al. (2004).

Imageability. Imageability refers to the ease of generating an image when a given word is presented (e.g., *comb* is higher in imageability than *caste*). We used the 7-point ratings collected by Cortese and Fugett (2004), which can be downloaded from http://myweb.unomaha.edu/~mcortese/norms%20link.htm (verified September 30, 2009).

AoA. Age of acquisition refers to the age at which a word is learned. We used the 7-point ratings collected by Cortese and Khanna (2008). These are available on http://myweb.unomaha.edu/~mcortese/norms%20link.htm (verified September 30, 2009).

Dependent variables

Rather than collecting new lexical decision data, we contacted the authors of the original studies and asked them whether they still had the full dataset (we thank the authors for their generosity). For each study, we received item-level information on token durations, mean lexical decision latencies, and mean lexical decision accuracy. Unless indicated otherwise, RTs were calculated from stimulus onset.

Results

The predictive power of the different frequency estimates

Frequencies were log transformed, and we only used words that were present in both the Hoosier Mental Lexicon database and across the various frequency databases. Table 1 presents the percentages of variance explained by the various frequency measures in the performance measures of the three studies we included.

There are several noteworthy aspects in Table 1. First, mirroring the findings from visual word recognition, the subtitle frequency measure based on the number of films in which a word occurs ($SUBTL_{CD}$) generally outperforms the other frequency measures. Second, among all the written frequency measures, KF67 frequency is the worst. Surprisingly, HAL frequency is not doing well either, suggesting that Internet-based frequency estimates may not be a good indicator of spoken word use. To provide convergent validation, we tested the Google frequency measure and obtained similar results. Finally, the spoken frequency measures of CELEX and BNC fare quite badly as well, arguably because they are based on small corpora of non-spontaneous speech (correlations based on Pastizzo and Carbone (2007) or ANC_{spoken} frequencies were equally disappointing). The rank order of the frequency measures was largely the same for the perceptual identification tasks (Experiment 1) and the naming task (Experiment 3) used in Luce and Pisoni (1998; see below). Given that the $SUBTL_{CD}$ measure is unequivocally the best measure, both for auditory and visual word processing (Brysbaert & New, in press), we will limit our frequency-related analyses to this index in the remainder of the article.

Another interesting aspect of Table 1 is the relatively low proportions of variance explained by frequency. They hover around 10%, which is much lower than the estimates previously reported for visual lexical decision performance. To ensure that the lower percentages were not due to the small number of stimuli, we examined the proportions of variance that were explained by frequency if we replaced the auditory lexical decision data from the experiments by the visual lexical decision data from the Balota et al. (2004) monosyllabic megastudy (http://www.psych.wustl.edu/coglab/labpub.htm, retrieved on September 30,2009). These data are shown in Table 2.

As can be seen by comparing Table 1 with Table 2, the frequency effect in auditory lexical decision RTs is about one third as strong as that in visual lexical decision RTs. To further investigate the issue, we also calculated the percentages of variance accounted for in auditory word naming (Luce and Pisoni, 1998, Experiment 3) and visual word naming (Balota et al., 2004; http://www.psych.wustl.edu/coglab/labpub.htm). These data are shown in Table 3. From this table, it is looks like there is not much difference between the frequency effect in visual word naming and auditory word naming (taking into account that the accuracy data were close to ceiling level). The frequency effect in naming is much smaller than in lexical decision, also for the auditory modality. Interestingly, the frequency effect is not completely absent in auditory word naming when a good frequency measure is used rather than KF67 (the estimate of Luce & Pisoni, 1998). This is yet another reminder that quality differences between the various frequency measures do have implications for the type of theoretical conclusions researchers draw from their data (see also Zevin & Seidenberg, 2002). With the benefit of hindsight, it is plausible that Luce and Pisoni might have come to (slightly) different conclusions if they had access to a better frequency measure at the time.

Luce and Pisoni (1998) also provided subjective familiarity ratings for their words (taken from Nusbaum, Pisoni, & Davis, 1984). When we entered these together with $SUBTL_{CD}$ frequency in a regression analysis on the lexical decision times, $SUBTL_{CD}$ frequency explained most of the variance and familiarity accounted for an additional 3.4% of unique variance (bringing R^2 up to 13.5%). When the familiarity ratings were entered with KF67 frequency, they explained relatively more variance (8.8%) and KF67 frequency added another 3.0% (bringing R^2

to 11.8%). This replicates Brysbaert and Cortese's (submitted) observation that familiarity ratings are particularly needed when one has as an inferior frequency measure, which should be a further incentive for researchers to drop the KF67 frequency.

The relative weights of the different variables

Now that we know which frequency measure to use ($SUBTL_{CD}$), we can examine the importance of the different variables (see Table 4 for the correlations between the various predictors). In addition, we can directly compare effects in auditory lexical decision and in visual lexical decision.

To estimate the impact of each predictor, we calculated two indices: The percentage of variance accounted for when the variable was the *only* variable in the regression analysis and the increase in R^2 when the variable was *added to* the other variables (see Table 5). In most instances, the impact of a variable decreases when correlated variables have been entered before (which gives an idea of the unique variance accounted for by the variable). However, in some of our analyses, due to non-additive effects, the increase in R^2 when the variable was introduced after the others actually was slightly higher than the zero-order R^2. For comparison purposes, Table 5 also includes the results of the same analyses on the Balota et al. (2004) visual lexical decision data.

Three surprising observations came out of Table 5. First, AoA was a more important variable than word frequency. Given the importance attached to the frequency effect (see above), this is noteworthy finding, as it suggests that a large part of the frequency effect is due to differences in AoA (see Smith et al., 2006, and Turner et al., 2004, for similar findings). A comparison with the visual lexical decision data confirms that the AoA effect is more important in the auditory modality than in the visual modality. In visual lexical decision (RTs), even after controlling for AoA and all other correlated variables, word frequency still accounted for 11% of the variance if the analysis was based on all words in the three studies, 12% more variance if the analysis was limited to the Luce and Pisoni words, 9% for the Cleland et al. words, and 16% for the Goh et al. words (see also Butler & Hains, 1979, and Morrison & Ellis, 1995 for other data on visual lexical decision).

A second surprising finding is the size of the stimulus duration effect on response times. Across studies, this is by far the most important predictor of response times, consistent with Goldinger’s (1996) warning that stimuli must be controlled for token duration.

The third surprising finding was the small size of the neighborhood density effect (except for the Goh et al. study), given the importance attached to this variable in the literature. One reason for this might be that neighborhood density *per se* is not an optimal estimator of competitive processes during word recognition. For example, Luce and Pisoni (1998) showed that a frequency-weighted neighborhood probability measure did much better than the simple neighborhood density measure for predicting perceptual identification performance.

The effects of uniqueness point and sound-spelling rime consistency were small as well and did not generalize across datasets. Given that most uniqueness points in the stimulus sets used coincided with the word end, it must be kept in mind that the range of this variable was very limited and, hence, the impact of the variable is likely to be underestimated in the analyses we have conducted. The absence of a consistency effect in lexical decision is also not entirely surprising, given that this effect tends to be larger in tasks such as speeded naming, where there is an emphasis on the production of phonology (Balota et al., 2004; Jared, McRae, and Seidenberg, 1990). To ensure that these patterns were not due to small sample size or restriction of range, we examined the proportions of variance that were explained by consistency if we replaced the auditory lexical decision data from the experiments by the visual *naming* data from Balota et al.'s (2004) monosyllabic megastudy. For the Luce and Pisoni words, consistency explained 0.8% of the variance in accuracy and 2.3% of the variance in RTs; for the Cleland et al. words, the percentages were respectively 1.4% and 1.5%; and for the Goh et al. words, they were 1.0% and 3.8%. Collectively, these results support the idea that consistency plays a very modest role in auditory lexical decision performance, at least in English. Finally, the effect of word imageability on auditory lexical decision is very modest as well.

Replacing time from stimulus onset by time from stimulus offset

To circumvent the problem of differences in stimulus durations, Luce and Pisoni (1998) started the time measurement at the *offset* of the stimulus token rather than at the onset. To evaluate to what extent this is a solution for the stimulus duration effect observed in Table 5, we ran a further set of regression analyses on the RTs minus the stimulus duration (see Table 6).

As can be seen in Table 6, although subtracting the stimulus duration from the response time ameliorated the confounding to a large extent in the Luce and Pisoni study, it actually *reversed* the problem in the other two studies (and also to some extent in Luce and Pisoni): RTs were now faster to words that took a long time to pronounce, compared to words that were pronounced faster. So, using total word duration as an estimate of the uniqueness point is not always a good strategy (as cautioned by Goldinger, 1996), because words may be recognized before the end of the final phoneme, either because their uniqueness point is earlier, or because the final phoneme is long.

Given the strong effect of stimulus duration on auditory lexical times, both when times measurement starts at word onset and when time measurement starts at word offset (in the reverse direction), we decided to examine in greater detail the effect of word onsets on auditory lexical decision performance. This will be addressed in the next section.

Do onset characteristics account for variance above and beyond lexical and semantic variables?

To further investigate the influence of stimulus duration on auditory lexical decision times, we decided to examine whether the token onset produced an effect *above and beyond* the token duration. There are two ways to study the possible effects of onset characteristics on auditory lexical decision times. The first is to repeat what has been done in word naming tasks. Here, researchers (e.g., Balota et al., 2004; Chateau & Jared, 2003; Treiman et al., 1995) have defined the first phoneme using a combination of dichotomous variables that encode the articulatory features of the sound (e.g., affricative, alveolar, bilabial, dental, …) and examined how much variance these variables explain.

Another approach is to look in detail at the actual speech signal. Such a study has been reported by Rastle et al. (2005), who carried out a fine-grained examination of onset effects in articulation. Using a delayed naming paradigm, they examined the temporal characteristics of the motor execution stage of speech production and documented two negatively correlated effects. First, speech onsets differ on their *execution-acoustic interval* (EAI), the interval between the signal to initiate the motor execution and the onset of acoustic energy. Second, they also differ in their acoustic *onset duration* (OD), the interval between the acoustic onset of a syllable and the acoustic onset of its vowel. For instance, Rastle et al. showed that it takes participants on average 228 ms to initiate the /s/ sound and 161 ms to pronounce it. In contrast, participants require 303 ms to initiate the /b/ sound, and only 16 ms to pronounce it. For auditory word recognition experiments, the onset duration variable is of particular importance, because the timer starts to run when the pronunciation of the word starts. What Rastle et al.'s analysis shows is that it takes (approximately) 145 ms longer before a listener hears the second phoneme in *sat* than in *bat*. This may have an impact on the time participants require to indicate that the stimulus is a word.

To examine whether word onset characteristics have an effect on auditory LDTs beyond word duration, we ran two additional analyses. In the first one, we used the 13 dummy variables proposed by Balota et al. (2004). In the second analysis, we used the onset times calculated by Rastle et al. (2005). Rastle et al. only analyzed consonant onsets, but this was no limitation given that only a very small minority of the stimuli in the studies under investigation started with a vowel. In the first and second steps, we entered the lexical and semantic variables of Table 4. One potentially controversial decision we made here was to include AoA in the second step (semantic variables). If it had been entered in the first step, its influence of course would have been much larger, at the expense of the frequency effect. In Step 3, we either entered the dummy variables or the onset durations. Table 7 shows the results.

The results from these regression analyses are broadly compatible with the earlier analyses. There were significant effects of frequency and AoA, neighborhood density was significant in two of three datasets (Cleland et al. and Goh et al.), and there were no clear effects of uniqueness point, consistency, or imageability across all datasets.

The interesting new finding is that onset characteristics explained a significant proportion of variance in RTs, above and beyond token duration and all the variables mentioned above. It varies from 3% in Luce and Pisoni (1998), 4% in Goh et al. (in press), to almost 18% in Cleland et al. (2006). Indeed, onset characteristics appear to account for more variance than most of the popular variables in auditory word recognition research.

Discussion

In this paper, we examined the impact of various factors thought to be important in auditory lexical decision performance. This was done by conducting multiple regression analyses on well-established datasets from different research groups, so that any observed results were unlikely to be due to the idiosyncrasies of a particular group. Some quite unexpected findings emerged from these analyses.

The first surprising finding was that word frequency was *not* the best predictor of auditory lexical decision performance. Auditory lexical decision frequency effects were about three times smaller than visual lexical decision frequency effects (compare Tables 1 and 2), although they were still larger than visual word naming frequency effects (compare Tables 1 and 3). More intriguingly, it seems likely that the bulk of the frequency effect is actually an AoA effect in disguise. AoA was the most robust factor in all our analysis and survived the partialling out of all lexical variables (see Table 7), whereas frequency accounted for very little variance once AoA was partialled out (see Tables 5 and 6). AoA has not been taken very seriously in the auditory word recognition literature, even though there have been studies pointing to its primacy over word frequency (Smith et al., 2006; Turner et al., 2004). Our analyses clearly show that the unique effect of AoA is considerably stronger in auditory lexical decision than in visual lexical decision (see Table 5).

Why is the AoA effect more influential in the auditory modality? Turner and colleagues suggested that there are separate lexicons associated with visual and spoken word recognition respectively, and that each lexicon is differentially affected by AoA. Specifically, early in life, language learning is mediated exclusively by the auditory modality, and it is thus likely that early acquired words have more impact on the organization of the phonological lexicon than on the organization of the orthographic lexicon. Some evidence for this hypothesis comes from brain imaging studies. Fiebach, Friederici, Müller, von Cramon, and Hernandez (2003) reported an increased activation in the *auditory* cortex when participants were reading *visually presented* early acquired words than when they were reading late acquired words.

A second surprising finding was the large effect of stimulus duration on lexical decision times. Although researchers have been warned repeatedly that they should control their stimuli for stimulus duration (e.g., Goldinger, 1996; Wurm et al., 2004), the magnitude of duration effects is nevertheless very sobering, in particular because there were virtually no differences in the number of phonemes between the stimuli we tested (88% consisted of 3 phonemes, 12% of 4 phonemes). This implies that the number of phonemes is not a good variable for estimating the stimulus duration of short, monosyllabic words. One could make the same case for the definition of the uniqueness point. There is little gain in controlling short stimuli for the *number* of phonemes up to the uniqueness point, given that phonemes can differ a lot in duration. Instead, it may make more sense to control for the *duration* up to the uniqueness point.

The analysis with the Rastle et al. (2005) measures indicates that the onset duration is an important predictor of auditory lexical decision times (remaining significant even when total stimulus duration is partialled out). As hypothesized, the execution-acoustic interval (EAI), which plays an important role in word naming studies, is not relevant in auditory LDTs, because timing only begins after the end of this interval. However, the time needed to pronounce the first consonant(s) substantially contributes to the time participants require before they can identify an auditorily presented word. To further clarify the issue, we examined the mean lexical decision residual times for the different phonemes, after controlling for word frequency, neighborhood density, uniqueness point, token duration, consistency, imageability, AoA, and number of phonemes (where applicable).

Table 8 presents these mean residual times as a function of dataset and phoneme. There are interesting trends that seem to hold across the datasets. For example, words beginning with stops and affricates (generally) produce the shortest times, while words beginning with fricatives produce the longest times. Words beginning with nasals fall in between. Obviously, there are exceptions to this general rule, and the rank orderings across the three datasets are also not identical. These variations are likely driven by differences in the regional accent of the speaker and other indexical properties. Just as voice keys take longer to detect onsets with low acoustic energy (e.g., /s/; Rastle & Davis, 2002), the human auditory perceptual system may also struggle more with some phonemes, hence delaying recognition times. Importantly, these results, coupled with the regression analyses, provide convergent evidence that some onsets provide information faster than others, and this then creates RT variance that is independent of the major psycholinguistic variables controlled thus far. Most researchers already know that onsets should be matched in experimental paradigms which measure the time taken to *produce* a vocal response to a word. However, the present findings indicate that onset characteristics also influence the time taken to *recognize* a word, suggesting that onset matching needs to be carried out even in the auditory lexical decision task. More in general, it looks like authors should present more information about the duration of the various segments of their stimulus words than they currently do.

Our analyses further revealed that auditory lexical decision performance is relatively little affected by neighborhood density, consistency, and imageability (unique percentages of variance accounted for less than 3%). We hasten to point out, however, that we do not consider this as evidence *against* a significant contribution of these variables. As Sibley et al. (2009) indicated, multiple regression analyses may not be the most powerful technique to examine the significance of a variable, especially if an unselected stimulus sample limits the range of the variable. Indeed, other authors have reported reliable effects of these variables (Luce & Pisoni, 1998, Experiment 1; Wurm et al., 2004; Ziegler et al., 2008). What our analyses do show, however, is that the unique variance accounted for by these variables, beyond that collectively accounted for by stimulus duration (or duration to the uniqueness point), onset characteristics, AoA, and frequency, is likely to be low. This agrees with findings from the visual modality, where Baayen et al. (2006) reported that the lower bounds for the percentage of variance explained by significant predictors was lower than 1% for all the non-frequency related variables (e.g., inflectional entropy, derivational entropy, word category, mean bigram frequency, and consistency). These low percentages mean that authors should worry less about perfectly matching on these variables when they design an auditory lexical decision study. It is much more important to look at the big four, of which two - AoA and onset characteristics - have not figured prominently thus far.

Last but not least, our study is a wake-up call that researchers should be more discriminating in the use of their frequency measure. Too many studies still rely on frequency measures (e.g., KF67, spoken CELEX) that are clearly inferior to the best measures available. This makes it more difficult to find reliable frequency effects, and leads to deficient stimulus matching, which increases the likelihood of spurious effects of variables related to frequency. These are weaknesses that can easily be remedied by making use of more contemporary frequency measures that are based on sufficiently large corpora.

References

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

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., 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.

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

Brysbaert, M., & Cortese, M. J. (2009). *Do the effects of familiarity and age-of-acquisition survive better word frequency norms?* Manuscript submitted for publication.

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

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.

Chereau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*, 341-360.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception & Performance, 32*, 104-119.

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

Conley, P., Burgess, C., & Hage, D. (1999). Large-scale databases of proper names. *Behavior Research Methods, Instruments, & Computers, 31*, 215-219.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in auditory and visual word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 1084-1096.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 81-94.

Cortese M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology, 60*, 1072-1082.

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

Cortese, M. J., & Khanna, M. M. (2008). Age of acquisition ratings for 3000 monosyllabic words. *Behavior Research Methods, 40*, 791-794.

Cutler, A. (1981). Making up materials is a confounded nuisance: or Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10*, 65-70.

Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 249-283). Amsterdam: Academic Press.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics, *Behavior Research Methods, 37*, 65-70.

Davis, C. J., & Taft, M. (2005). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review, 12*, 904-910.

De Moor, W., & Brysbaert, M. (2000). Neighborhood-frequency effects when primes and targets are of different lengths. *Psychological Research, 63*, 59-62.

Fiebach, C. J., Friederici, A. D., Müller, K., von Cramon, D. Y., & Hernandez, A. (2003). Early and late learned words are represented differently in the brain. *NeuroImage, 19*, 1627-1637.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition, 28*, 1109-1115.

Goh, W. D., Suárez, L., Yap, M. J., & Tan, S. H. (in press). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review*.

Goldinger, S. D. (1996). Auditory lexical decision. *Language & Cognitive Processes, 11*, 559-567.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language, 28*, 501-518.

Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language & Cognitive Processes, 8*, 379-411.

Humphreys, L. G. (1978). Research on individual differences requires correlational analysis, not ANOVA. *Intelligence, 2*, 1-5.

Ide, N., & Macleod, C. (2001). The American National Corpus: A Standardized Resource of American English. *Proceedings of Corpus Linguistics 2001*, Lancaster UK.

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

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear & Hearing, 23*, 2-40.

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

Kiyonaga, K., Grainger, J., Midgley, K., & Holcomb, P. J. (2007). Masked cross-modal repetition priming: An event-related potential investigation. *Language & Cognitive Processes, 22*, 337-376.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Lemhoefer, K., Dijkstra, A., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 12-31.

Lewis M. B. & Vladeanu, M. (2006). What do we know about psycholinguistic effects? *Quarterly Journal of Experimental Psychology, 59*, 977-986.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Doctoral dissertation, Indiana University, Bloomington, IN.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech perception* (pp. 122-147). Cambridge, MA: MIT Press.

Luce, P., & Pisoni, D. B. (1998). Recognizing spoken words. The neighborhood activation model. *Ear & Hearing, 19*, 1-36.

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.

MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. Psychological Review, 89, 483-506.

MacKay, D. G. (1987). The organization of perception and action: A theory for language and other cognitive skills. New York: Springer.

Marslen-Wilson, W. D. (1987). Parallel processing in spoken word recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical Access and Representation* (pp. 3–24). Cambridge, MA: Bradford.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148-172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113*, 181-190.

Meunier, F., & Segui, J. (1999). Frequency effects in auditory word recognition. *Journal of Memory & Language, 41*, 327-344.

Miller, K. M., & Swick, D. (2003). Orthography influences the perception of speech in alexic patients. *Journal of Cognitive Neuroscience, 15*, 981-990.

Morrison, C. M., & Ellis, A. W. (1995). The roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 116-133.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76,* 165-178.

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.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*. Bloomington, IN: Speech Research Laboratory, Psychology Department, Indiana University.

Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality & Social Psychology, 12,* 194-199.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*, 220-244.

Pastizzo M. J., & Carbone, R. F. (2007). Spoken word frequency counts based on 1.6 million words in American English. *Behavior Research Methods, 39*, 1025-1028.

Pattamadilok, C., Morais, J., Ventura, P., & Kolinsky, R. (2007). The locus of the orthographic consistency effect in auditory word recognition: Further evidence from French. *Language & Cognitive Processes, 22*, 1-27.

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

Rastle, K., & Davis, M. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception & Performance, 28*, 307-314.

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception & Performance, 31*, 1083-1095.

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study. *Bulletin of the Psychonomic Society, 27,* 489.

Shaoul, C., & Westbury, C. (2009) A USENET corpus (2005-2009) Edmonton, AB: University of Alberta (downloaded from http://www.psych.ualberta.ca/~westburylab/downloads/usenetcorpus.download.html)

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (in press). Learning orthographic and phonological representations in models of monosyllabic and bisyllabic naming. *European Journal of Cognitive Psychology*.

Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, *14*, 230-237.

Slowiaczek, L. M., Soltano, E. G., Wieting, S. J., & Bishop, K. L. (2003). An investigation of phonology and orthography in spoken-word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *56A*, 233-262.

Smith, P. T., Turner, J. E., Brown, P. A., & Henry, L. A. (2006). The distinct contributions of age of acquisition and word frequency in auditory word perception. *Quarterly Journal of Experimental Psychology*, *59*, 2121-2134.

Soares, I., Collet, L., & Duclaux, R. (1991). Electrophysiological correlates of auditory lexical decision: An attempt to test the “cohort model”. *International Journal of Neuroscience*, *57*, 111-122.

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.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition*, *22*, 259-282.

Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory & Language, 58*, 366-379.

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.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition, 26,* 1282-1291.

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. *Psychonomic Bulletin & Review, 7*, 320-326.

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

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

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory & Language, 40*, 374-408.

Vitevitch, M. S. (2002). The influence of onset-density on spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 28*, 270-278.

Whatmough, C., Arguin, M., & Bub, D. (1999). Cross-modal priming evidence for phonology-to-orthography activation in visual word recognition, *Brain & Language, 66*, 275-293.

Wurm, L. H., Vakoch, D. A., & Seaman, S. R. (2004). Recognition of spoken words: Semantic effects in lexical access. *Language and Speech, 47,* 175-204.

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

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

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in reading and other tasks. *Journal of Memory & Language, 47*, 1-29.

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

Ziegler, J. C., & Muneaux, M. (2007). Orthographic facilitation and phonological inhibition in spoken word recognition: A developmental study. *Psychonomic Bulletin & Review, 14,* 75-80.

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

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighbourhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory & Language, 48*, 779-793.

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 & Cognition, 34*, 643-661.

Acknowledgements

Melvin Yap and Marc Brysbaert contributed equally to this project. We thank Alexandra Cleland, Micah Geer, Winston Goh, and Paul Luce for generously providing us with their data. We are also grateful to Kathy Rastle and Arthur Samuel for their very useful input during the preparation of this manuscript.

Footnotes

1. Goldinger's list also included priming effects, which are beyond the scope of this paper.
2. Subsequent analyses confirm that none of the claims made in this paper depend on the specific density measure used. The same findings are obtained with density A. The reason why one density measure was dropped was because of the very high correlation between both measures ($r > .90$), which created a collinearity problem in the regression analyses.
3. We thank the authors who kindly provided us with these data.
4. Although this corpus is presented as having this size, in reality the useful information is limited to some 500 billion words, still much more than any other corpus.
5. KF67 frequencies can easily be obtained from 3 different sources: Elexicon (http://elexicon.wustl.edu/), the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm), and N-Watch (Davis, 2005; http://www.pc.rhul.ac.uk/staff/c.davis/Utilities/). An intriguing (and worrying) observation is that the three sources do not always correspond in their counts. For instance, for the word *aged*, Elexicon and MRC give a KF67 frequency of 18, whereas N-Watch gives a frequency of 27. In contrast, the word *ants* is not present in MRC (suggesting a frequency of 0), whereas Elexicon and N-Watch list a frequency of 7. A likely explanation for these differences is that at some point in time some lemmatization took place, which got integrated in some "KF67" files and not in others.

Table 1. Proportion of variance explained in auditory lexical decision performance by the different frequency measures. KF = Kučera & Francis (1967) frequency norms; Celex = the Center for Lexical Information word-form frequency norms (Baayen et al., 1993); Zeno = the Zeno et al. (1995) frequency norms; HAL = the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996); SBT_{WF}= SUBTL wordform frequency norms (Brysbaert & New, in press); SBT_{CD} = SUBTL contextual diversity frequency norms (Brysbaert & New, in press); Cel_{sp} = the Center for Lexical Information spoken frequency norms (Baayen et al., 1993); BNC_{sp} = the British National Corpus (BNC) spoken frequency norms (Leech et al., 2001).

	KF	Celex	Zeno	HAL	SBT_{WF}	SBT_{CD}	Cel_{sp}	BNC_{sp}
Luce & Pisoni Acc (n=726)	6.4	9.4	12.3	6.5	8.7	11.0	5.6	5.9
Luce & Pisoni RT (n=726)	5.6	6.9	8.8	7.0	9.4	10.4	5.8	7.5
Cleland et al. Acc (n=353)	11.0	14.6	14.8	14.5	15.4	16.1	10.2	12.5
Cleland et al. RT (n=353)	6.0	7.5	7.6	7.6	9.4	9.0	6.5	8.2
Goh et al. Acc (n=184)	6.1	7.0	12.4	6.3	9.7	11.8	4.5	9.1
Goh et al. RT (n=184)	5.9	8.6	10.5	6.8	13.3	13.9	5.8	8.9
Mean	6.8	9.0	11.0	8.1	11.0	12.0	6.4	8.7

Table 2. Proportion of variance explained in visual lexical decision task (VLDT) performance by the different frequency measures, based on Balota et al. (2004) words overlapping with words in the auditory datasets. KF = Kučera & Francis (1967) frequency norms; Celex = the Center for Lexical Information word-form frequency norms (Baayen et al., 1993); Zeno = the Zeno et al. (1995) frequency norms; HAL = the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996); SBT_{WF}= SUBTL wordform frequency norms (Brysbaert & New, in press); SBT_{CD} = SUBTL contextual diversity frequency norms (Brysbaert & New, in press); Cel_{sp} = the Center for Lexical Information spoken frequency norms (Baayen et al., 1993); BNC_{sp} = the British National Corpus (BNC) spoken frequency norms (Leech et al., 2001).

	KF	Celex	Zeno	HAL	SBT_{WF}	SBT_{CD}	Cel_{sp}	BNC_{sp}
Luce & Pisoni Acc (n=673)	9.2	13.2	12.7	9.5	10.9	15.1	7.7	8.8
Luce & Pisoni RT (n=673)	23.8	29.4	31.1	24.8	31.7	38.6	20.4	25.1
Cleland et al. Acc (n=287)	31.1	35.0	34.7	30.3	32.5	35.8	27.4	32.9
Cleland et al. RT (n=287)	46.2	50.3	50.6	48.3	51.7	54.6	40.6	49.1
Goh et al. Acc (n=177)	9.4	8.2	11.4	4.4	8.3	9.1	5.0	8.4
Goh et al. RT (n=177)	26.2	27.4	27.4	24.4	36.0	38.6	21.3	33.8
Mean	24.3	27.2	28.0	23.6	28.5	31.9	20.4	26.4

Table 3. Proportion of variance explained in auditory and visual naming performance by the different frequency measures. The visual naming analyses are based on Balota et al. (2004) words overlapping with words in the auditory dataset. KF = Kučera & Francis (1967) frequency norms; Celex = the Center for Lexical Information word-form frequency norms (Baayen et al., 1993); Zeno = the Zeno et al. (1995) frequency norms; HAL = the Hyperspace Analogue to Language (HAL) frequency norms (Lund & Burgess, 1996); SBT_{WF}= SUBTL wordform frequency norms (Brysbaert & New, in press); SBT_{CD} = SUBTL contextual diversity frequency norms (Brysbaert & New, in press); Cel_{sp} = the Center for Lexical Information spoken frequency norms (Baayen et al., 1993); BNC_{sp} = the British National Corpus (BNC) spoken frequency norms (Leech et al., 2001); ANT = auditory naming task; VNT = visual naming task.

	KF	Celex	Zeno	HAL	SBT_{WF}	SBT_{CD}	Cel_{sp}	BNC_{sp}
Luce & Pisoni ANT Acc (n=860)	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Luce & Pisoni ANT RT (n=860)	0.6	1.5	2.0	1.8	3.2	3.3	1.2	2.0
Balota et al. VNT Acc (n=803)	0.2	0.9	1.2	1.4	1.2	1.2	0.5	0.9
Balota et al. VNT RT (n=803)	1.1	2.0	2.9	3.0	3.4	3.7	2.0	3.0

Table 4. The correlations between word frequency, neighborhood density, uniqueness point, feedback rime consistency, imageability, and age of acquisition. The matrix below the diagonal is based on the subset of words (n = 954) for which we have values for all six variables. The matrix above the diagonal is based on the full dataset (n = 1090) and each correlation is computed using the maximum number of observations available.

	1	2	3	4	5	6
1. Log $SUBTL_{CD}$ Frequency	-	.096**	.118***	.018	.049	-.748***
2. Neighborhood Density B	.066*	-	.022	-.086**	.080**	-.119***
3. Uniqueness Point	.149***	.014	-	.015	.019	-.048
4. Feedback Rime Consistency	.030	-.088**	.010	-	.087**	-.096**
5. Imageability	-.100**	.063†	.014	.088**	-	-.394***
6. Age of Acquisition	-.699***	-.084**	-.063†	.095**	-.301***	-

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

Table 5. The proportion of variance explained in auditory and visual lexical decision performance by different lexical variables, when response times are measured from stimulus onset. The visual lexical decision analyses are based on Balota et al. (2004) words yoked to the words in the respective auditory dataset. The first number gives the percentage of variance explained when the variable is entered alone; the last number gives the percentage of variance the variable adds when it is introduced in addition to the other six variables. Freq = Word frequency; DensB = Density B; Uniq = Uniqueness point; Sdur = Stimulus duration; Con = Feedback rime consistency; Image = Imageability; AoA = Age of acquisition; VLDT = Visual lexical decision task.

	Freq	DensB	Uniq	Sdur	Con	Image	AoA
Luce & Pisoni Acc (n = 665)	6*** - <1*	<1† - 0	0 - 0	<1* - 0	0 - 0	3*** - 1**	11*** - 2***
Luce & Pisoni RT (n = 665)	10*** - 1***	3*** - 0	<1** - 0	20*** - 13***	0 - 0	3*** - <1**	15*** - 1***
VLDT Acc (n = 665)	15*** - 4***	0 - 0	0 - 0	NA	0 - 0	3*** - 2***	18*** - 1***
VLDT RT (n = 665)	39*** - 12***	0 - 0	<1† - 0	NA	0 - <1**	3*** - 3***	36*** - 2***
Cleland et al. Acc (n = 287)	10*** - 3**	0 - 1†	<1 - 1†	2* - 3**	0 - 0	<1 - 0	8*** - 0
Cleland et al. RT (n = 287)	7*** - 0	0 - 2**	2* - 1*	23*** -23***	0 - 0	<1 - 0	11*** - 2**
VLDT Acc (n = 287)	35*** - 5***	0 - 0	<1 - 0	NA	0 - 0	4*** - <1*	35*** - 1*
VLDT RT (n = 287)	56*** - 9***	2** - 0	0 - 0	NA	0 - 0	4*** - <1*	51*** - 1**
Goh et al. Acc (n = 176)	11*** - 0	0 - 0	0 - 0	<1 - 0	0 - <1	4**- 0	29*** - 14***
Goh et al. RT (n = 176)	13*** - <1	5** - 3**	0 - 0	25*** - 13***	10*** - <1	4** - 0	22*** - 5***
VLDT Acc (n = 176)	9*** - 2*	0 - 0	0 - 0	NA	0 - 0	8*** - 3*	15*** - 1
VLDT RT (n = 176)	38*** - 16***	0 - 1*	<1 - 0	NA	4** - 0	4** - 2*	27*** - 0

*** $p < .001$, ** $p < .01$; * $p < .05$; † $p < .10$, in a hierarchical multiple regression involving all seven variables

Table 6. The proportion of variance explained in auditory lexical decision performance by different lexical variables, when response times are measured from stimulus offset. The first number gives the percentage of variance explained when the variable is entered alone; the last number gives the percentage of variance the variable adds when it is introduced in addition to the other six variables. Freq = Word frequency; DensB = Density B; Uniq = Uniqueness point; Sdur = Stimulus duration; Con = Feedback rime consistency; Image = Imageability; AoA = Age of acquisition.

	Freq	DensB	Uniq	Sdur	Con	Image	AoA
Luce & Pisoni RT (n = 665)	8*** - 1***	0 - 0	0 - 0	3*** - 5***[A]	0 - 0	1** - <1**	10*** - 2***
Cleland et al. RT (n = 287)	10*** - 0	1* - 1**	0 - <1*	43*** - 39***[A]	0 - 0	0 - 0	9*** - 2**
Goh et al. RT (n = 176)	8*** - < 1	1 - 3**	1 - 0	17*** - 22***[A]	0 - <1	<1 - 0	15*** - 6***

*** p < .001, ** p < .01; * p < .05; † p < .10, in a hierarchical multiple regression involving all seven variables
[A] Effect is in the opposite direction: faster responses for longer duration words

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

Predictor Variables	Luce & Pisoni		Cleland et al.		Goh et al.	
	LDT RT (*n* = 665)	LDT Accuracy (*n* = 665)	LDT RT (*n* = 276)	LDT Accuracy (*n* = 276)	LDT RT (*n* = 175)	LDT Accuracy (*n* = 175)
Step 1: Standard Lexical Variables						
Word frequency	-.264***	.244***	-.349***	.338***	-.329***	.360***
Neighborhood density B	.011	.027	.173**	-.131†	.203**	-.081
Uniqueness Point	-.035	.010	.090	-.079	.019	-.036
Stimulus Duration	.415***	-.068†	.505***	-.172**	.415***	-.067
Consistency	.003	-.051	-.031	.011	-.094	-.063
Number of phonemes	NA	NA	.072	-.032	NA	NA
Adj. R^2	**.262*** **	**.067*** **	**.345*** **	**.113*** **	**.383*** **	**.103*** **
Step 2: Semantic Variables						
Imageability	-.105**	.133**	.042	.028	.010	-.058
Age of acquisition	.189***	-.198***	.296**	-.033	.353***	-.587***
Adj. R^2	**.306*** **	**.123*** **	**.367*** **	**.113*** **	**.451*** **	**.282*** **
	ΔR^2 = .044* **	**ΔR^2 = .056*** **	**ΔR^2 = .022** **	**ΔR^2 = .000**	**ΔR^2 = .068*** **	**ΔR^2 = .054*** **
Step 3a: Onset Features						
Affricative	-.008	-.059	-.355***	-.019	-.085	-.145
Alveolar	-.031	-.327	.000	.044	.071	-.017
Bilabial	-.165	-.297†	-.159**	.050	.045	-.195*
Dental	-.111	-.120	-.123**	-.036	.070	-.008
Fricative	-.040	-.299	NA	NA	NA	NA
Glottal	-.107	-.294	-.066	-.036	.004	.078
Labiodental	.056	-.321	-.103†	.004	-.003	.001
Liquid	-.144	-.195	-.262***	-.007	-.098	.092
Nasal	NA	NA	NA	NA	NA	NA
Palatal	.341***	-.093	.372***	-.045	.283*	-.180
Stop	.112**	.003	.003	.031	.043	-.011
Velar	.014	.037	.203***	-.135	.035	-.089
Voiced	.059	.069	.280***	.007	.067	.072
Adj. R^2	**.333*** **	**.147*** **	**.543*** **	**.113*** **	**.493*** **	**.362*** **
	ΔR^2 = .027* **	**ΔR^2 = .024** **	**ΔR^2 = .176*** **	**ΔR^2 = .000**	**ΔR^2 = .042* **	**ΔR^2 = .054** **
Step 3b: Onset Temporal Properties						
Execution-acoustic interval	.055	.021	-.007	.110	-.005	.261*
Onset duration	.196**	-.022	.398***	.067	.199†	.195
Adj. R^2	**.322*** **	**.123*** **	**.511*** **	**.113*** **	**.481*** **	**.291*** **
	ΔR^2 = .016* **	**ΔR^2 = .000**	**ΔR^2 = .144*** **	**ΔR^2 = .000**	**ΔR^2 = .030** **	**ΔR^2 = .009**

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

Table 8. Mean residual lexical decision RTs as a function of dataset and onset.

Luce & Pisoni		Cleland et al.		Goh et al.	
Phoneme	Residual RT (ms)	Phoneme	Residual RT (ms)	Phoneme	Residual RT (ms)
p	-36.74	g	-113.29	d	-40.02
n	-36.67	b	-83.71	b	-23.68
r	-29.45	v	-35.27	g	-23.14
l	-23.90	k	-28.61	t	-20.75
b	-19.26	d	-22.98	k	-16.79
j	-15.50	h	-21.11	m	-13.31
ð	-14.38	r	-6.91	n	-11.75
d	-10.58	l	-6.44	w	-11.59
f	-5.18	t	-6.17	l	-11.20
θ	-2.70	tʃ	10.18	dʒ	-6.38
v	1.01	w	17.71	r	-3.18
g	1.95	n	22.60	h	12.09
m	7.42	p	23.42	f	23.82
t	9.14	m	24.14	p	26.46
w	17.07	f	35.59	s	53.07
h	18.12	ʃ	37.59	v	93.85
k	19.30	s	99.02		
dʒ	26.13				
ʃ	31.66				
s	55.22				

Mean residual RTs are reported only for onsets with at least 5 observations in a dataset.