Vowel Duration Affects Visual Word Identification: Evidence That the Mediating Phonology Is Phonetically Informed

Georgije Lukatela, Thomas Eaton, Laura Sabadini, and M. T. Turvey
University of Connecticut and Haskins Laboratories

What form is the lexical phonology that gives rise to phonological effects in visual lexical decision? The authors explored the hypothesis that beyond phonological contrasts the physical phonetic details of words are included. Three experiments using lexical decision and 1 using naming compared processing times for printed words (e.g., *plead* and *pleat*) that differ, when spoken, in vowel length and overall duration. Latencies were longer for long-vowel words than for short-vowel words in lexical decision but not in naming. Further, lexical decision on long-vowel words benefited more from identity priming than lexical decision on short-vowel words, suggesting that representations of long-vowel words achieve activation thresholds more slowly. The discussion focused on phonetically informed phonologies, particularly gestural phonology and its potential for understanding reading acquisition and performance.

There is a hypothesis of long standing that reading involves "inner speech." Essentially, the hypothesis claims that reading a word engages, at some point, a representation that is not substantially distinct in formal structure from the word's pronunciation and/or acoustic form. As Huey (1908/1968, p. 122) remarked, "The simple fact is that the inner saying or hearing of what is read seems to be the core of ordinary reading, the 'thing in itself,' so far as there is such a part of such a complex process."

The classical hypothesis of inner speech seems far removed from modern thinking on the nature of underlying representations for spoken and written words. The prevalent view in linguistic theory is that a word's lexical representation does not encode its physical phonetic properties, especially the temporal. A word's representation is an abstraction away from the dynamics of the word's articulatory form. In linear generative phonology, for example, the phonological form of any given entry in the internal lexicon is conveyed by a sequence of consonants and vowels represented as columns of feature values (e.g., Kenstowicz & Kisserberth, 1979). The features in these distinct columns represent attributes of segmental articulation and acoustics. The abstract manner in which they do so, however, produces a phonological code in the lexicon that is at some remove from how a given word is realized in articulation and sound.

It is illuminating to consider generative linear phonology as the interpretative context for the recent observation that a word such as *plead* is responded to more slowly in the visual lexical decision task than a word such as *pleat* (Abramson & Goldinger, 1997).

Georgije Lukatela, Thomas Eaton, Laura Sabadini, and M. T. Turvey, Department of Psychology, University of Connecticut, and Haskins Laboratories, New Haven, Connecticut.

Correspondence concerning this article should be addressed to Georgije Lukatela, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695. E-mail: lukatela@uconnvm.uconn.edu

Although a consensus view on the details of visual lexical decision has yet to be achieved, it is commonly assumed that it is a kind of retrieval or recognition. Essentially, the participant asks him- or herself the question Does the optical form I'm looking at match any of the forms in my memory for words? A word's form in memory could be closely related to its optical form. That is, it could be an orthographic form. If so, then matching would be conducted, presumably, in terms of visual features or in terms of the letters they compose. Alternatively, or additionally, a word's form in memory could be phonological in the (abstractly specified) generative linear sense. If so, then matching would be preceded by a conversion (e.g., by grapheme-to-phoneme rules) of the optical form into a phonological form, and the matching itself would be conducted, presumably, in terms of abstract phonological segments.

It is not obvious how an account of a visual lexical decision that presumes only orthographic lexical representations could address the observation that *plead* is identified as a word more slowly than *pleat*. The two words are identical in number of letters and alike in frequency of occurrence and number of orthographic neighbors—three dimensions that might be expected to affect the time required to find a match between a printed word and its orthographic representation in memory. It is similarly not obvious, and for much the same reasons, how the difference in latency between *pleat* and *plead* could be addressed by an account of visual lexical decision that assumes abstract phonological lexical representations as defined by generative linear phonology. The two words are identical in the number of phonemic segments and alike in frequency of occurrence and number of phonological neighbors.

Pressing the phonological view, we can note that for both *plead* and *pleat*, the phoneme /i/ corresponding to the letters *ea* in the two words would be represented in linear phonology by one and the same feature list. The spatial and temporal aspects of pronouncing *ea* in *plead* and *pleat* are not the same, however, and neither are their corresponding sound patterns. Typically, vowels are longer in duration when followed by a voiced consonant such as /d/ than when followed by a voiceless consonant such as /t/ (e.g., Chen, 1970; Klatt, 1973; Peterson & Lehiste, 1960; Port, 1981; Stevens

& House, 1963).[1] The effect of this vowel lengthening on spoken-word duration is partially but not fully compensated by the shorter duration of the voiced consonant relative to the unvoiced consonant (see Figures 1 and 2 in Port, 1981; A. Abramson, personal communication, August 2002; L. Lisker, personal communication, August 2002). With monosyllabic words such as *plead* and *pleat,* the lengthening of the vowel by a voiced consonant is simultaneously a lengthening of the embedding word. The overall phonetic realization is longer.

The aforementioned internal duration changes are the consequences of the dynamics of articulation—the time evolution of constrictions, the phasing of gestures. The contrast between *plead* and *pleat* in visual lexical decision mirrors, therefore, the physical facts of pronunciation and the corresponding acoustics—facts that are not available in the underlying linear phonology but are available, in principle, in the surface form derived by rule from the abstract underlying phonological form. Perhaps it should be argued that visual lexical decision, when expressed through the formalisms of generative linear phonology, is understood as a decision at the level of the derived surface forms rather than at the level of the base underlying forms. That is to say, the decision that a letter string is a word is more aptly termed *postlexical* rather than *lexical.*

To retain the conventional lexical perspective on visual lexical decision, it would be necessary to show that the phonetics–phonology divide has been too sharply drawn—that phonetic details, temporal included, inhere in underlying phonological patterns. There have been several recent efforts to reduce the gap between phonetics and phonology (e.g., Boersma, 1998; Flemming, 1995; Jun, 1995; Kirchner, 2001; Mohanan, 1993; Steriade, 1997). Typifying these efforts is the observation that phonological patterns are crucially affected by the quality and duration of phonetic features (Steriade, 1997). Most noteworthy and most thoroughgoing among these efforts is gestural phonology (Browman & Goldstein, 1990, 1992, 1995; Gafos, 2002). In contrast to more conventional theories, the primitives of gestural phonology are not consonants and vowels, not phonemes, and not features. Rather, the primitives are gestures and constellations of gestures of the vocal tract. In approximate terms, a *gesture* consists of "characteristic patterns of movement of vocal tract articulators" (Browman & Goldstein, 1986, p. 223), and a *gestural constellation* is a small number of potentially overlapping gestures that compose an utterance.

Because a word's representation in gestural phonology is composed of dynamical systems (functional synergies) and their phase relations, it incorporates information about the temporal coordination of the gestures that express consonants and vowels (Saltzman & Munhall, 1989). It is important for the classical hypothesis of inner speech that a gestural constellation capture both the properties unique to a word and the properties of the word that are common to multiple words. That is, under this formulation, a word's entry in the internal lexicon is close to its pronunciation and to its corresponding acoustic form. Idiosyncratic and systematic properties of a word are not held separate from the word (as in generative phonology) but are part and parcel of its unitary representation.

The upshot of the preceding discussion is that the classical hypothesis of inner speech in reading (e.g., Egger, 1904; Huey, 1908/1968) can be given a modern interpretation in terms of representations that integrate phonology and phonetics. Reading a word engages a phonological form that represents not only pho-

nologically significant distinctions of traditional concern but also physical phonetic details. It inculcates category (or discrete) and gradient (or continuous) distinctions alike.

Abramson and Goldinger's (1997) research, as suggested above, exploited the fact that vowels in words spoken at a conversational pace are typically lengthened when followed by voiced consonants (e.g., /b/, /d/, /m/) relative to voiceless consonants (e.g., /k/, /p/, /t/). The vowel in *plead* is lengthened relative to the same vowel in *pleat,* and the time required to utter *plead* is longer than the time required to utter *pleat.* Within gestural phonology, this duration difference (i.e., this phonetic length effect) would be understood as a consequence of different patterns of phasing of consonant and vowel gestures. In a general effort to close the phonetics–phonology gap, this phonetic length effect would be understood, in less specific terms, as the incorporation by phonology of information about physical phonetic details. One might expect, therefore, that if the lexical forms mediating visual word recognition integrate "cognitive" phonology and "physical" phonetics, then written words that differ in vowel length (when spoken) might also differ in their associated latencies of visual lexical decision. In particular, one might expect that these visual decision latencies are longer for phonetically longer words.

Experiment 1: Replication of Abramson and Goldinger (1997)

Experiment 1 was a visual lexical decision experiment that replicated the design of Abramson and Goldinger's (1997) experiments but with a more comprehensively controlled set of stimuli. At issue was whether lexical decision times would be longer for monosyllabic words such as *plead, pile, roam,* and *graze* (in which the final consonant is voiced) than for monosyllabic words of similar consonant–vowel composition, frequency, and neighborhood size such as *pleat, pipe, roach,* and *grate* (in which the final consonant is unvoiced).

Method

Participants. Twenty-four undergraduates at the University of Connecticut participated in partial fulfillment of course requirements in introductory psychology. (All participants in the present and subsequent experiments were informed as to the nature of the experiment, and all gave their consent to participate.)

Materials. Fifty-six monosyllabic words constituted the test set. Approximately 25% of these words were borrowed from Abramson and Goldinger (1997; compare their appendix with Table 1). Each word was three, four, or five letters long. The structure of each word was CVC, CVCC, CCVC, or CCVCV (C = consonant, V = vowel). The intermediate vowel was followed by a voiced consonant (e.g., *plead*) in 28 words and by a voiceless consonant (e.g., *pleat*) in the remaining 28 words, yielding equal numbers of words with long and short vowels.

Each short-vowel word had a counterpart in a long-vowel word producing a matched pair. The two matched words shared the initial CV or CCV pattern (e.g., *pleat* and *plead, pipe* and *pile, roach* and *roam, grate* and *graze;* see Table 1). The mean frequency of the short-vowel words was

[1] Laeufer (1992) provided an overview of the effects of obstruent voicing on vowel duration in various linguistic and extralinguistic (e.g., speaking rate) contexts. The implication is that the effect requires a general phonetic account in terms of gestural timing (rather than the invocation of language-specific phonological rules).

Table 1

Target Words and Corresponding Mean Lexical Decision Times in Experiment 1

Short vowel		Long vowel	
Word	Reaction time (ms)	Word	Reaction time (ms)
bats	493	bags	523
brake	525	braid	560
clap	537	clan	524
crate	578	crane	559
creep	531	creed	567
grate	587	graze	592
guts	567	gums	604
hats	507	hang	572
lake	539	lane	581
leap	546	lean	555
lick	537	lint	561
lope	596	lobe	592
mates	525	manes	507
pipe	519	pile	525
pits	518	pins	568
pleat	513	plead	547
roach	558	roam	635
ropes	548	robes	589
sap	532	sag	509
slap	594	slab	573
slip	505	slid	585
snack	525	snag	534
spit	564	spin	561
spike	545	spine	557
swap	542	swan	575
swipe	516	swine	569
taps	546	tags	529
trick	514	trim	581

7.93, and the mean frequency of the long-vowel words was 8.60, according to Kučera and Francis (1967). (On the presumption that differences would be more easily detected in slower responding, all words were chosen to be of relatively low frequency.) The mean neighborhood size was 8.21 for the short-vowel words and 7.75 for the long-vowel words. Neither the word frequency difference nor the neighborhood size difference was statistically significant (both $Fs < 1$).

In addition to the 56 test words, there were 77 ordinary nonwords, 48 pseudohomophones, and 69 filler words, for a total of 250 stimuli. The highly diversified experimental stimuli were intended to minimize the development of any specific response strategy. A set of 56 practice stimuli was composed similarly to the set of 250 experimental stimuli. All stimuli in the experimental and practice sets were orthographically legal and pronounceable as single syllables.

Design. Each participant saw all 250 stimuli, with the order of presentation varied across participants. The stimuli were divided into three approximately equal subsets defining three blocks of trials, with a brief rest after each block.

Procedure. Participants were tested one at a time. Each participant sat in front of the monitor of a DIGITAL 466 computer in a well-lit room. The viewing distance was about 60 cm. The refresh rate of the VENTURIX monitor was 70 Hz, making a refresh cycle (i.e., a "tick") of 14.3 ms.

Each trial in the practice and experimental series consisted of a sequence of two visual events in the same location on the center of the screen. The first event was a row of five uppercase Xs (XXXXX) presented for 30 ticks (430 ms), and the second event, which immediately followed the first, was a target stimulus in lowercase letters presented for 60 ticks (860 ms). The interval between trials within a block was 1 s.

Presentation and control of stimuli were through DMASTR software (developed at Monash University and the University of Arizona by K. I.

Forster & J. C. Forster, 1990). Participants were told that in each trial there would be a sequence of two stimuli. They were instructed to decide as quickly and as accurately as possible whether the second stimulus, a lowercase letter string, was a word or a nonword. They responded by pressing one key for yes and another key for no. If the decision latency exceeded 1,800 ms, "RESPOND FASTER!" appeared on the screen. Further, if the participant hit the wrong key, "WRONG" appeared on the screen. (The procedure for Experiment 1 was approved by the Institutional Review Board of the University of Connecticut, as were the similar procedures for Experiments 2–4.)

Results

All decision latencies were trimmed minimally by applying a 100-ms cutoff for fast responses and an 1,800-ms cutoff for slow responses. The outliers constituted less than 0.5% of all responses (see criteria for truncation suggested by Ulrich & Miller, 1994, p. 69). The same trimming procedure was applied to all experiments reported in the present article.

The mean reaction time and mean error were 541 ms and 6.11%, respectively, for the short-vowel words (e.g., *pleat, pipe, roach,* and *grate*) and 564 ms and 7.00%, respectively, for the long-vowel words (e.g., *plead, pile, roam,* and *graze*). The mean correct decision latencies for each of the 28 short-vowel words and 28 long-vowel words are reported in Table 1. Because the stimuli were matched, a direct comparison of the matched stimuli by means of a paired *t* test on the correct decision latencies was conducted. The mean difference of 23 ms favoring short-vowel words was significant, $t(27) = 3.88, p < .001$. As inspection of Table 1 reveals, the short-vowel word was responded to faster on average in 20 of the 28 pairs. An analysis of variance (ANOVA) corroborated the finding of a main effect of vowel length in both the subjects analysis, $F_1(1, 23) = 12.63, p < .002$, and the items analysis, $F_2(1, 54) = 14.65, p < .001$. The error analysis was insignificant (both $Fs < 1$).

The results of Experiment 1 are in agreement with those of Abramson and Goldinger (1997). With largely different instances of short- and long-vowel words and a refinement in selection that equated the two kinds of words in both frequency and neighborhood size, we found that lexical decision was slower for long-vowel words than for short-vowel words.

Experiment 2: Identity Priming as a Function of Vowel Length

The conjecture that phonetic information about a word's physical time evolution is available to lexical phonological representations suggests that the distinction between short-vowel and long-vowel words can be interpreted as a distinction between the activation rates of their respective representations. Specifically, the activation rate for short-vowel words is higher than that for long-vowel words. This particular interpretation of the outcome of Experiment 1 can be put to the following test: If the two classes of words differ in the rates at which they become activated, then they should differ in the degree to which they benefit from identity priming. Presumably, with other things being equal, a lexical item that reaches threshold more quickly will benefit less from an identity prime than will a lexical item that reaches threshold more slowly.

In Experiment 2, we examined lexical decision on a target as a function of identity priming within a mask–prime–mask–target

sequence (Lukatela, Frost, & Turvey, 1999; Lukatela & Turvey, 1994) at two onset asynchronies between prime and target. We compared identity prime–target sequences of the kind *pleat–pleat* and *plead–plead* relative to nonidentity prime–target sequences of the kind *plead–pleat* and *pleat–plead*. The primes in the nonidentity sequences were words of opposite vowel length that were orthographically and phonologically similar, but semantically dissimilar, to the targets.

Method

Participants. Sixty-eight undergraduates at the University of Connecticut participated in partial fulfillment of course requirements in introductory psychology. A participant was assigned to one of two groups within one of two stimulus-onset asynchrony (SOA) conditions, with 17 participants per group.

Materials. The base set of 56 words from Experiment 1 was used to construct 112 different prime–target pairs. Four kinds of pairs were assembled, with the prime in uppercase and the target in lowercase: (a) Each short-vowel word was preceded by itself to create 28 identity pairs (e.g., PLEAT–*pleat*), (b) each long-vowel word was preceded by itself to create 28 identity pairs (e.g., PLEAD–*plead*), (c) each short-vowel word was preceded by a long-vowel word to create 28 nonidentity pairs (e.g., PLEAD–*pleat*), and (d) each long-vowel word was preceded by a short-vowel word to create 28 nonidentity pairs (e.g., PLEAT–*plead*).

There were, in addition, 192 filler pairs. The filler pairs were highly diversified to make the development of any specific response strategy unlikely. They consisted of 24 word–word pairs with the prime and target related associatively, 24 word–word pairs with the prime and target not related associatively, 24 nonword–word pairs with rhyming prime and target, 24 nonword–word pairs with the prime and target sharing the two initial letters, 24 unrelated nonword–nonword pairs, 24 nonword–nonword pairs of rhyming prime and target, 24 pseudohomophone–homophone pairs, and 24 pseudohomophone–ordinary nonword (nonhomophone) pairs. The nonwords, as either primes or targets, were orthographically legal and pronounceable as single syllables. Separate from the preceding stimulus pairs, a further 56 pairs were generated for practice trials.

Design. The major constraint was that a given participant never encountered a given prime–target pair more than once. This was achieved by dividing the participants in each SOA condition into two groups, A and B. Each participant in Group A saw one half of the pairs from the identity test list and one half of the pairs from the nonidentity test list, and all filler pairs. Each participant in Group B saw the other halves of the test lists together with all filler pairs. More specifically, in this design, Group A experienced the identity prime–target pairs (e.g., PLEAD–*plead*, CLAP–*clap*) and the nonidentity prime–target pairs (e.g., LAKE–*lane*, SNACK–*snag*) that corresponded, respectively, to Group B's nonidentity prime–target pairs (PLEAD–*pleat*, CLAP–*clan*) and identity prime–target pairs (LAKE–*lake*, SNAG–*snag*). Given this design, it was meaningless to compare performances on identity and nonidentity pairs within a group and it was meaningless to compare the two groups with respect to the magnitude of the identity–nonidentity difference. For the preceding reasons, interactions involving group were ignored in the 2 × 2 × 2 (Group × Vowel Length × Prime–Target Similarity) ANOVA used to evaluate the experimental results.

In sum, within an experimental sequence each participant saw a total of 250 stimulus pairs, of which 56 pairs were test pairs. As in Experiment 1, the experimental sequence was divided into three approximately equal subsets, with a brief rest after each.

Procedure. The essential conditions of presentation were those of Experiment 1. Several additions and modifications were necessary to implement the four-field masked priming methodology.

Each trial consisted of a sequence of four visual events at the same location at the center of the screen. The sequence began with a pattern mask consisting of a row of eight number signs (########) presented for

38 ticks or monitor refresh cycles (resulting in a display duration of 543 ms). This initial mask was followed immediately (interstimulus interval of 0 ms) by the prime stimulus in uppercase letters presented for either 1 tick (14.3 ms) or 9 ticks (129 ms). The prime was followed immediately by a so-called intervening mask composed of a row of seven ampersands (&&&&&&&). This intervening mask was presented for 3 ticks (42.9 ms) in the case of the 1-tick prime and for 8 ticks (114 ms) in the case of the 9-ticks prime. A target in lowercase followed the intervening mask immediately to yield an SOA of 57 ms when the prime was 1 tick and an SOA of 243 ms when the prime was 9 ticks. The target was presented for 38 ticks (543 ms).

Participants in the shorter SOA condition were not told of the presence of the primes. Participants in the longer SOA condition were told that they would see two letter strings but were only required to make a lexical decision on the second, lowercase letter string.

Results and Discussion

Analyses were performed on the correct latencies and on the errors associated with the targets in the identity and nonidentity conditions for short-vowel and long-vowel targets. The mean values for the four subsets at each of the two prime–target onset asynchronies are summarized in Table 2.

ANOVAs on the latency and error measure were conducted with factors of group, onset asynchrony, identity, and vowel length. Independent of onset asynchrony, the benefit of identity priming on latency of lexical decision was greater for long-vowel targets (50 ms) than for short-vowel targets (23 ms), $F_1(1, 64) = 11.11$, $p < .01$, and $F_2(1, 52) = 8.84, p < .01$, respectively. The single main effect was that lexical decisions were slower on long-vowel targets (589 ms) than on short-vowel targets (562 ms), $F_1(1, 64) = 56.52, p < .001$, and $F_2(1, 52) = 9.36, p < .01$, respectively. In the error analysis, the benefit of identity priming was greater for long-vowel targets (6.0%) than for short-vowel targets (1.2%), $F_1(1, 64) = 8.50, p < .01$, and $F_2(1, 52) = 5.07, p < .05$, respectively, independent of onset asynchrony. The advantage for short-vowel targets over long-vowel targets was greater at the onset asynchrony of 57 ms (6.0% fewer errors) than at the onset asynchrony of 243 ms (.56% fewer errors), $F_1(1, 64) = 8.80, p < .01$, and $F_2(1, 52) = 9.00, p < .01$, respectively.

In sum, the results corroborate and extend the vowel length effect first discovered by Abramson and Goldinger (1997) and

Table 2

Mean Lexical Decision Times (in Milliseconds) and Percentage Error (in Parentheses) as a Function of Prime-Target SOA, Prime Type, and Vowel Length of Target in Experiment 2

	Target	
SOA and prime	Short	Long
57 ms		
Identity	551	563
	(9.4)	(14.1)
Control	571	610
	(9.2)	(17.9)
243 ms		
Identity	550	565
	(7.1)	(5.2)
Control	575	615
	(9.7)	(12.7)

Note. SOA = stimulus onset asynchrony.

confirmed by Experiment 1. Long-vowel words differ from short-vowel words in two respects: (a) They are recognized more slowly in the visual lexical-decision task and (b) they benefit more from visual identity priming.

The second difference was the major focus of Experiment 2. It warrants further consideration. Specifically, there is the question of whether the two versions of the nonidentity prime–target sequences, *pleat–plead* and *plead–pleat*, were functionally equivalent. Where prime and target are similar in their phonological compositions but not identical, and where the time to process the prime is severely curtailed, competition may arise between the phonological activity engendered by the prime and that engendered by the target. Experiments on masked priming with rhyming words have found that whereas *nose* primes *nose* positively, *rose* and *rows* prime *nose* negatively (Lukatela & Turvey, 1996). With temporal restrictions on prime processing, a short-vowel prime such as *pleat* might be more completely processed within the available time than a long-vowel prime such as *plead*. That is, within the mask–prime–mask–target sequence, *pleat* could create more competition for *plead* than vice versa. If so, then the results of Table 3 could reflect differences in the nonidentity sequences more than differences between long- and short-vowel targets in susceptibility to identity priming.

We conducted a subsidiary experiment (Experiment 2A) with the nonidentity sequences (e.g., *pleat–plead* and *plead–pleat*) of Experiment 2 replaced by new nonidentity sequences (e.g., *roach–plead* and *roam–pleat*, respectively). In these new nonidentity sequences, the primes differed from the targets in vowel length and in orthographic and phonological composition, thereby eliminating the grounds for the kind of prime–target competition envisaged by Lukatela and Turvey (1996). There were 20 participants. All were tested in the 57-ms onset asynchrony condition. The results shown in Table 3 replicated those of the main experiment shown in Table 2. Lexical decisions were slower for long-vowel targets (584 ms) than for short-vowel targets (557 ms), $F_1(1, 18) = 14.52, p < .001$, and $F_2(1, 52) = 5.64, p < .02$, respectively, and long-vowel targets seemed to benefit more from identity priming (40 ms) than did short-vowel targets (8 ms), $F_1(1, 18) = 4.25, p < .05$, and $F_2(1, 52) = 3.46, p = .07$, respectively. An additional subsidiary experiment (Experiment 2B) with 42 participants was directed at the equivalency of the new nonidentity sequences. It compared

them (e.g., *roach–plead* and *roam–pleat*) with control sequences (e.g., XXXX–*plead* and XXXX–*pleat*, respectively). The results are shown in Table 3. There was no interaction between prime type and vowel length of the target ($Fs < 1$), meaning that the effect of a long-vowel prime on a short-vowel target was the same as that of a short-vowel prime on a long-vowel target. The outcomes were simply faster responses to short-vowel targets (by 27 ms), $F_1(1, 40) = 22.95, p < .0001$, and $F_2(1, 52) = 6.41, p < .01$, respectively, and faster responses with XXXX priming (by 10 ms), $F_1(1, 40) = 6.37, p < .01$, and $F_2(1, 52) = 4.30, p < .05$, respectively. On the basis of the subsidiary experiments, it seems safe to conclude that long-vowel words benefit more from visual identity priming than do short-vowel words.

Experiment 3: Vowel-Length Effect With a Larger Stimulus Set and Articulatory Suppression

The results of Experiment 1 and Experiment 2 (including Experiments 2A and 2B) confirmed Abramson and Goldinger's (1997) finding that words with long vowels are responded to more slowly in the visual lexical decision task than words with short vowels. The wide-ranging theoretical significance of the finding is recommendation in itself for circumspection. Word frequency in the preceding experiments of the present article was controlled through word selection by means of the Kučera and Francis (1967) frequency norms derived from the Brown corpus of approximately one million items. The latter norms, however, are not ideal. According to Kučera and Francis (1967) *boxer, icing,* and *joker* have the same frequency count as *ioire, gnome,* and *assay,* contrary to the reality of daily usage and experience (Gernsbacher, 1984). It would be prudent, therefore, to assess the vowel-length effect in an experiment in which frequency is controlled through more reliable standards of common, modern-day usage. For this purpose, the CELEX corpus (Baayen, Piepenbrock, & Gulikers, 1995) of approximately 18 million items was used in Experiment 3. The CELEX measures were buttressed in subsequent post hoc analyses by the subjective frequency estimates for 2,398 monosyllabic English words (Balota, Pilotti, & Cortese, 2001).

An additional reason for reevaluating the vowel-length effect is that across the voiced–unvoiced stimulus pairs in Experiments 1 and 2 other distinctions prevailed. For example, the postvocalic consonants differed within a pair, albeit unsystematically, in place and manner of articulation. Patently, all such contrasts cannot be fully controlled in the context of necessary controls for onset, vowel, frequency of occurrence, and neighborhood size. A pragmatic response to this problem of possible confounds due to uncontrolled phonological differences is to increase the number of pairs explicitly controlled on the dimension of voicing. In Experiment 3, the number of stimulus pairs composing a vowel-length condition was more than double that of Experiment 1 (66 pairs vs. 28 pairs).

The purpose of Experiment 3 was twofold. As previewed above, it was conducted primarily to evaluate the latency difference favoring words with short vowels with a larger and better controlled stimulus set. In addition, Experiment 3 was conducted to assess the persistence of the short-vowel advantage under conditions of brief exposure during concurrent articulatory suppression. Supporters of the classical view of inner speech often appealed to auditory imagery. Specifically, they observed that one is frequently aware during silent reading of hearing a voice in one's

Table 3

Mean Lexical Decision Times (in Milliseconds) and Percentage Error (in Parentheses) as a Function of Prime Type and Vowel Length of Target in Experiments 2A and 2B

Experiment and prime	Target	
	Short	Long
2A		
Identity	553	564
	(8.2)	(8.2)
Nonidentity	560	604
	(9.6)	(14.9)
2B		
Nonidentity	585	614
	(9.0)	(13.8)
XXXX	577	602
	(9.7)	(11.9)

head (Huey, 1908/1968). Subsequently, dismissal of the necessity of inner speech to reading followed simply from noting the many counterinstances in which reading occurred in the absence of an inner voice. From a modern perspective, the issue of mediation by a "voice in the head" can be addressed by the use of techniques that minimize the opportunity to pronounce a word silently, such as curtailment of visual processing time and suppression of articulation.

Method

Participants. Twenty-five undergraduates at the University of Connecticut participated in partial fulfillment of course requirements in introductory psychology.

Materials. The base set comprised 132 monosyllabic words. For the new set of 66 short-vowel words, the mean CELEX frequency (per million) was 8.26, and the mean CELEX frequency (per million) for the 66 long-vowel words was 9.3. A concerted effort was made to select those pairs for which any small difference in CELEX frequency favored the long-vowel member of the pair, $t(65) = -1.05$, $p < .05$. For later comparisons, we also determined the mean Kučera and Francis (1967) frequency: 8.34 for short, 9.42 for long, $t(65) = -1.05$, $p < .05$. The Kučera and Francis mean frequencies agreed with the CELEX mean frequencies. The mean neighborhood size was 10.65 for the short-vowel words and 9.98 for the long-vowel words ($t < 1$); the mean familiarity was 3.63 for short-vowel words and 3.53 for long-vowel words ($t < 1$). Other remarks and explanations concerning materials and design that were given in Experiment 1 remained valid in Experiment 3.

Procedure. Several modifications were made to the procedure used in Experiment 1. Each trial consisted of a sequence of three visual events on the same location in the center of the screen. Event 1 was a row of seven hash marks (#######) presented for 40 ticks (572 ms); Event 2 was an immediately following target stimulus in lowercase letters presented for 15 ticks (215 ms); Event 3 was an immediately following display of six uppercase Xs (XXXXXX) presented for 40 ticks (572 ms). The intertrial interval was 100 ticks (1,430 ms). Event 1 signaled that the participant should begin speaking softly and repeating the sequence of numbers 1 through 5 at a constant speed of approximately four digits per second (a simple task that has been shown to successfully suppress articulation: Baddeley, Lewis, & Vallar, 1984). Event 2 signaled lexical decision, and Event 3 signaled the end of the trial. During the practice trials, participants received feedback on the appropriateness of the pace at which they performed the number-articulation task.

Results and Discussion

The subjects' mean reaction time and the mean error were 572 ms and 14.67% for the long-vowel words and 544 ms and 12.6% for the short-vowel words, respectively. The mean correct decision latencies for each of the 66 short-vowel words and 66 long-vowel words are reported in Table 4. Because the stimuli were matched, a direct comparison of the matched stimuli by means of a paired t test on the correct decision latencies was conducted. The mean items' difference of 27 ms favoring short-vowel words was significant, $t(65) = 3.52$, $p < .001$.[2] As inspection of Table 4 reveals, the short-vowel word was responded to faster on average in 45 of the 66 pairs. ANOVA on latency revealed a significant vowel-length effect, $F_1(1, 24) = 48.02$, $p < .0001$; $F_2(1, 130) = 11.79$, $p < .001$. The corresponding error ANOVA was nonsignificant.

A multiple regression analysis on latency was also performed with five independent variables. The quantitative independent variables CELEX frequency (CFR) and Kučera and Francis frequency (KFFR) were logarithmically transformed. The quantitative inde-

pendent variables familiarity (FAM) and neighborhood size (N) and the dummy variable vowel length (VL)—with categories long and short—were not transformed. Vowel length was the most significant regressor, $t(126) = 4.43$, $p < .0001$. Also significant were familiarity, $t(126) = -3.48$, $p < .001$; logCFR, $t(126) = -3.36$, $p < .001$; and logN, $t = 2.03$, $p < .05$.

The above treatment of dummy-variable regression assumes that the individual regressions for the five independent variables are parallel. If these regressions are not parallel, then the dummy regressor interacts with one or more of the quantitative regressors. To test the null hypothesis of no interaction between VL and FAM, and between VL and logCFR, we dichotomized the quantitative independent variables FAM and logCFR into two categories, median high and median low. ANOVAs (2 × 2) substantiated the null hypothesis for VL with FAM, $F(1, 128) = 1.14$, $p > .05$, and for VL with logCFR, $F(1, 128) < 1$. Other regressors being of less practical significance, we can conclude that the preceding multiple regression yielded correct estimates about the contribution of the individual independent variables to lexical decision latency. We can also conclude that the vowel-length effect persisted in Experiment 3 despite the limited opportunity to verbalize the words at presentation. The outcome is consistent with the general observation that articulatory suppression does not interfere with the derivation of a phonological code from print (Besner, 1987).

In the stimulus set of the present experiment, there were 46 tense vowel pairs and 20 lax-vowel pairs. A separate evaluation of the vowel-length effect for these tense and lax vowels is important for both theoretical and experimental reasons. Theoretically, the distinction is drawn in linear and nonlinear phonologies between lax and tense vowels as underlyingly *short* and *long*, respectively, and experimentally it is known that voicing-induced vowel lengthening holds for both types (Port, 1981). From the perspective of deriving a surface form from an underlying form, the lengthening of a lax vowel is a simpler proposition (formally speaking) than the lengthening of a tense vowel (see the Appendix). Simple t tests on the lax-vowel pairs and tense-vowel pairs of the present experiment revealed comparable vowel-length effects of 38 ms, $t(19) = 2.63$, $p < .02$, and 22 ms, $t(45) = 2.54$, $p < .02$, respectively.

In summary, the results of Experiments 1–3 confirm Abramson and Goldinger's (1997) finding that phonetically longer words are responded to more slowly in the visual lexical decision task than phonetically shorter words. First and foremost, this demonstration

---

[2] As a check on further potential sources of contamination, six pairs involving homophones and three pairs involving words of possibly questionable classification were deleted from the list. The mean difference between long- and short-vowel words for the remaining 57 pairs was 29 ms, $t(56) = 3.54$, $p < .001$. An additional, similar analysis was conducted in which the three pairs of questionable classification were eliminated together with two pairs (*bile–bikes, nib–nip*) in which the long-vowel word was less familiar than the short-vowel word. The latter two pairs yielded the largest long-vowel versus short-vowel differences of 242 ms and 203 ms. In the absence of the aforementioned five pairs, the mean difference between long- and short-vowel words for the remaining 61 pairs was 22 ms, $t(60) = 3.38$, $p < .001$. With the questionable five pairs excluded, the regression analysis yielded vowel length, $t(116) = 4.23$, $p < .0001$; familiarity, $t(116) = -3.28$, $p < .001$; and logCFR, $t(116) = -3.40$, $p < .001$, as the only significant variables. Similarly, for the tense versus lax comparison, the vowel effect was 29 ms for lax vowels, $t(18) = 2.42$, $p < .03$, and 19 ms for tense vowels, $t(41) = 2.42$, $p < .02$.

Table 4

Target Words and Corresponding Lexical Decision Reaction Time (RT, in Milliseconds) in Experiment 3

Short vowel		Long vowel	
Word	RT	Word	RT
bat	510	*bags*	536
bikes	486	*bile*	728
brake	502	*braid*	634
caves	573	*cage*	518
cape	538	*cane*	566
clap	535	*clan*	544
clot	625	*clog*	626
cope	535	*code*	524
crap	531	*crab*	519
crave	543	*craze*	499
creep	528	*creed*	564
dip	490	*digs*	551
dope	527	*dole*	604
dove	559	*dose*	635
fats	547	*fads*	623
fake	565	*fame*	517
gates	586	*gaze*	553
goats	514	*goals*	552
grate	589	*graze*	561
grips	569	*grim*	576
gut	557	*gum*	558
hats	501	*ham*	575
heap	527	*heel*	581
hush	564	*hum*	613
jeep	486	*jeans*	545
kits	629	*kin*	643
lake	542	*lane*	551
leap	561	*lean*	547
lip	495	*lid*	590
lice	697	*lime*	585
lope	554	*lobe*	585
lust	537	*lung*	548
mats	564	*mam*	593
mite	611	*mime*	608
moat	565	*moan*	622
nip	497	*nib*	700
nut	525	*nun*	649
pipe	523	*pile*	548
pit	537	*pin*	525
pike	603	*pine*	558
plait	620	*plaid*	574
plush	578	*plum*	553
rape	506	*rage*	543
rat	501	*ram*	550
rip	518	*rig*	585
ropes	539	*robe*	572
sap	625	*sag*	628
skip	507	*skid*	594
slap	520	*slab*	573
slit	541	*slid*	570
slips	502	*slim*	546
slut	579	*slum*	493
snack	490	*snag*	526
spate	458	*spade*	570
spat	588	*span*	539
spike	525	*spine*	580
swap	562	*swan*	549
swipe	552	*swine*	628
tat	668	*tab*	619
taps	535	*tan*	543
tip	517	*tin*	582
tope	633	*toad*	581
trash	502	*tram*	612
trips	534	*trim*	601
wake	577	*wage*	562
wits	610	*wig*	561

of a vowel-length effect in visual lexical decision is compelling evidence for the prominence (and, perhaps, primacy) of phonology in visual word recognition. Such an effect does not follow from the dual-route cascade model (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) with standard parameters. Additionally, such an effect does not follow straightforwardly from major distributed processing models (e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996). If it were the case that letters representing voiced consonants map to their corresponding phonemes with weaker strengths than letters representing voiceless consonants, then it might be possible for models of the latter kind to produce a *plead* versus *pleat* difference. Only appropriate simulations could confirm or reject this possibility. A favorable outcome would not be without problems, however. The question would then have to be raised as to why the statistics of letters to phonemes would differentiate voiced and voiceless in the aforementioned manner.

From a purely experimental point of view, a common stance is that evidence for phonological involvement in visual word recognition ought to be in the form of an influence on positive lexical decisions (e.g., Davelaar, Coltheart, Besner, & Jonasson, 1978). It is typically observed in English that a putative phonological manipulation—regularity of a word's grapheme-to-phoneme correspondence—affects naming latency but does not affect positive lexical decision latency (e.g., Hino & Lupker, 1996; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Waters & Seidenberg, 1985). In Serbo-Croatian, by way of contrast, phonological influences on positive lexical decision are readily demonstrated through manipulations of letters that receive different phoneme assignments in the independent but partially overlapping Roman and Cyrillic alphabets (see Lukatela & Turvey, 1998, for a review). We underscore that the results of Experiments 1–3 were obtained with the positive lexical decision latencies. By the commonly held criterion, the result is evidence of phonological involvement in English visual word recognition.

For present purposes, the main value of the demonstration of a vowel-length–phonetic-length effect in visual lexical decision lies in its relevance to the hypothesis that lexical representations integrate phonology and phonetics. Lukatela, Eaton, Lee, and Turvey (2001) reported evidence of subphonemic influences on visual word recognition that was potentially explainable by either a static feature-based or a dynamic gesture-based account. They asked whether the priming of a word (e.g., *sea, film, basic*) by a rhyming nonword would depend on the nonword's phonemic-feature similarity to the word. The question was asked within a mask–prime–target–mask sequence. Lukatela et al. found that nonword primes that differed from their targets by a single phonemic feature or a single gesture (initial voicing as in *zea, vilm, pasic*) led to faster target lexical decisions than nonword primes that differed by more than a single phonemic feature or by more than a single gesture (e.g., *vea, jilm, sasic*). The implication of the results of Experiments 1–3 is that this observed superiority of *zea* over *vea* as a prime for *sea* is better explained in terms of a notion that is closer to the temporally informative gestures of gestural phonology than the timeless features of generative linear phonology.

Experiment 4: The Vowel-Length Effect in Naming

The two primary responses used in investigations of visual word recognition are lexical decision and rapid naming. These two

responses are not constrained in the same way by the underlying memory stores and processes. This latter understanding derives from the effects that specific experimental manipulations have on the two responses. Here are a few examples. Associative priming effects are usually two to three times larger in lexical decision than in naming (Neely, 1991). Phonological priming effects are lexically dependent in lexical decision (word–word pairs show inhibition, nonword–word pairs show facilitation) and lexically independent in naming (both word–word and nonword–word pairs show facilitation and to the same degree; e.g., Lukatela, Carello, & Turvey, 1990). Irregular spelling–sound correspondences affect lexical decision less than they affect naming (e.g., Seidenberg et al., 1984). The major implication of these few examples is, for our present purposes, that lexical forms are reflected to a lesser degree in naming than in lexical decision. In Experiment 4, we wished to take advantage of this implication. If the vowel-length effect originates at the level of lexical representations, and if naming engages these representations to a lesser degree than lexical decision, then the vowel-length effect in naming should be less pronounced than the vowel-length effect in lexical decision. In Experiment 4, we tested this expectation of the vowel-length effect in the naming of visually presented words. The experiment was a replication of Experiment 3, with rapid naming replacing lexical decision as the response measure.

Method

Participants. Twenty-four undergraduates at the University of Connecticut participated in partial fulfillment of course requirements in introductory psychology.

Materials, design, and procedure. The materials and experimental design replicated those of Experiment 3. The additional task requirement of Experiment 3 used to suppress articulation was not part of Experiment 4.

Results and Discussion

The mean naming latency and the mean error were 539 ms and 0.9% for the short-vowel words and 548 ms and 1.2% for the long-vowel words, respectively. ANOVA revealed that the 9-ms vowel-length effect was significant by participants, $F_1(1, 23) = 25.55, p < .001$, but not by items, $F_2(1, 130) = 2.39, p > .10$. The error analysis found no hint of a vowel-length effect (both $Fs < 1$). Given the reproducibility of the vowel-length effect in Experiments 1–3 using the lexical decision task, we should take seriously the present result indicating a weak vowel-length effect in the naming task. A weakened effect in naming relative to its magnitude in lexical decision should be interpreted, therefore, as evidence for the effect's lexical origin—specifically, that the effect informs about the nature of phonological representations in the lexicon.

General Discussion

The hypothesis framing the present research was twofold: that representations in the lexicon mediate phonological effects in visual lexical decision and that these representations are simultaneously phonological and phonetic. Two kinds of evidence consistent with the hypothesis were reported. First, in Experiments 1–3, we confirmed and extended the finding of Abramson and Goldinger (1997) that visual lexical decision is sensitive to the vowel durations of the spoken words that the written words en-

code. Second, the interpretation of this vowel-length effect as originating at a lexical level was reinforced by Experiment 4. Using pronunciation latency as the measure of visual word processing, Experiment 4 found only a weak vowel-length effect in sharp contrast to the strong vowel-length effect found in the lexical decision task. This difference would be expected if performance in the naming task was less referential of lexical representations than performance in the lexical decision task, as is customarily argued.

The present results provide further support for the idea that visual word recognition is led by phonology. One variant of this idea is the *universal phonological principle* (Perfetti, Zhang, & Berent, 1992; Perfetti & Zhang, 1995); another is the *phonological coherence hypothesis* (Lukatela, Frost, & Turvey, 1999; Van Orden & Goldinger, 1994; Van Orden, Pennington, & Stone, 1990). What is presumed to be universal is the automatic, "reflexive" activation of a word's phonology in word identification for all writing systems and most words.

Common to the universality principle and the coherence hypothesis is the claim that the mapping is most systematic, most nearly one-to-one, between orthographic structure and phonological structure. Least systematic is the orthographic-to-semantic mapping and intermediate in systematic form is the mapping between phonological structure and semantic structure (but see Seidenberg, 1995). For the phonological coherence hypothesis, the advantage of a mapping that is nearly one-to-one is the rapid achievement of self-consistency within the matrix of connections linking the processing units of the two structures. The implied rapidly attained coherence of phonological codes puts them in the position of playing a key role in mediating the coherence and stability of other ongoing linguistic processes (Gotlob, Goldinger, Stone, & Van Orden, 1999; Van Orden & Goldinger, 1994; Van Orden et al., 1990). With respect to the timing of internal processes, the activation of a word's phonology occurs, according to both the principle and the hypothesis, at the earliest moment consistent with a given writing system's basic units. A number of recent investigations involving priming procedures with pattern-masked primes have supported the aforementioned hypothesis and principle. These investigations suggest that the lower limit on processing time in both naming and lexical decision is set by the time scale of resolving a word's phonology (e.g., Lukatela et al., 1999).

What Is the Form of the Mediating Phonology?

In the introduction, we suggested that traditional generative phonology using abstract representations could, in principle, address the Abramson and Goldinger (1997) vowel-length effect if the effect arose postlexically. Traditional generative phonology assumes an abstract input that does not formally encode properties such as universally noncontrastive vowel length. Such properties are thus defined as phonetic. Nonetheless, within the traditional method of representation, the derived surface phonology is less distant from the phonetic form and thus the rules of a derivation could include a rule to the effect: Lengthen the vowel in the context of a following voiced consonant.

Nonlinear autosegmental phonology. A robust account, however, requires that we be specific about the nature of the rule or constraint that lengthens the vowel in words such as /bɪd/. A rule of the form "V → long/___ voiced obstruent" is insufficient in the absence of a feature for length. We are encouraged, therefore, to explore the possibility of whether a sufficient rule might follow

from the formalism of autosegmental phonology (see the Appendix). Within this phonology are specific means to capture some forms of vowel duration–vowel length within lexical phonological representations. Application of the autosegmental formalism to the present problem reveals that the long-vowel versus short-vowel distinction due to voicing is not present in the underlying forms of monosyllabic words with either lax or tense vowels. It also reveals that the distinction cannot be obtained by derivation in a formally self-consistent manner (see the Appendix). That is, the distinction in question is also not available in the surface form, that which is closer to phonetics. This means that if access to postlexical representations yielded the Abramson and Goldinger (1997) vowel-length effect, then those representations could not have resulted from a derivation or constraint-set satisfaction using traditional input and features. Any adequately specific postlexical form must arise as the output of some much later phonetic module that would presumably encode properties such as noncontrastive vowel length. Such a module would necessarily be active in speech. Two possible expectations from the conjectured module would be a weak vowel-length effect under conditions of articulatory suppression (Experiment 3) and a strong vowel-length effect in the naming task (Experiment 4). Neither of the preceding expectations was evident in the data. A further argument that can be leveled against the addition of a late-acting phonetic module is that it would give rise to an (unnecessarily) inelegant framework (Kirchner, 2001).

Gestural phonology. It was suggested in the introduction that a more coherent account of the Abramson and Goldinger (1997) vowel-length effect in visual lexical decision would be provided by a system of phonology wedded to phonetic explanation at the outset (eliminating the need for a derivational process). As noted, the phonology most befitting the latter bill is gestural or articulatory phonology (Browman & Goldstein, 1986, 1989, 1990, 1992, 1995, 2001; Byrd, 1996; Fowler, 1995; Gafos, 2002; Saltzman & Byrd, 2000; Saltzman & Munhall, 1989). In gestural phonology, the low-dimensional cognitive phonology and the high-dimensional physical phonetics are complementary aspects of a single, complex dynamical system. Simply, there is no phonetic–phonological gap. Clements (1992) argued that replacing featural representation with gestural representation would result in too many degrees of freedom that would give rise to, among other things, unattested phonological contrasts and classes. Kirchner (2001) showed, however, that an optimality theoretic framework could constrain considerable phonetic detail in the input or underlying representation so as to maintain the appropriate number and degree of cognitive phonological contrasts and categories. Said differently, the enhancement of the underlying form with phonetic details, gestural or otherwise, need not result in excessively powerful phonological generalizations. To the contrary, it could allow for a necessarily specific form of lexical representation such as that implicated in the present experiments.

The Relation of Reading to Speech

The present experimental results have possibly broad implications depending on the extent that (a) the experimental results are reliable and (b) an interpretation of them in terms of gestural phonology proves to be reasonable. Specifically, they lend support to A. M. Liberman's (1992, 1996, 1998, 1999; I. Liberman, Shankweiler, & Liberman, 1989) self-described unconventional theory of speech and the theory of reading to which it gives rise. The

unconventional or motor theory of speech is intimately connected to gestural phonology. It takes the ultimate constituents of speech to be articulatory gestures rather than sounds, and it accords them the status of a natural class, a phonetic modality. Because of their inherently phonetic nature, gestures need no cognitive translation to render them suited to the requirements of linguistic processing. They have, thereby, a distinct advantage over the orthographic constituents of reading, however construed (e.g., graphemes; Rey, Ziegler, & Jacobs, 2000), for the orthographic constituents must be assigned their linguistic values through agreed-upon conventions. Gestures have other linguistic virtues. Under the management of a specialized phonetic module (A. M. Liberman & Mattingly, 1989), they can be overlapped and merged (i.e., coarticulated), making possible the rapid production and perception of the commutable units on which the productivity of language so heavily depends. In Liberman's unconventional theory, the gestures of the speaker are the objects of the listener's perception.

The unconventional motor theory of speech is of significance to the theory of reading on two counts. It identifies what is required to achieve reading success, and it identifies the major obstacles to that success. With respect to the requirement for success, the would-be reader must learn to map print directly to gestural phonology (and not to meaning as often presumed; A. M. Liberman, 1998, 1999; I. Liberman et al., 1989). Only phonology provides the reader with access to language's openness or productivity. That is, only through phonology can the reader exploit language's ability to encode an indefinitely large number of meanings. Simply put, the strategy for success is to enter the language mode and to enter it at the earliest level of its nested operations, namely, the phonological level.

With respect to the obstacles to success, they arise from demands of the alphabet principle that are at odds with the fluent operation of the phonetic module. To achieve mastery of the alphabet principle, the would-be reader must become aware of the phonemic composition of spoken words (where a phoneme may be considered as a particular cohering of gestures). Being able to identify letters and knowing their sounds is not sufficient. Unfortunately for the would-be reader, an explicit ability to break words apart into their phonemes is irrelevant to the efficient (automatic, nonconscious) function of the phonetic module. Consequently, the ability does not follow ordinarily from simply learning to speak. It typically requires special effort and particular phonological exercises.

The challenge of becoming phonologically aware arises from two of the previously mentioned aspects of gestures. The inherently phonetic nature of gestures means that they function outside the domain of language properties that normally attract attention. Gestures escape conscious notice because, unlike letters, for example, no cognitive translation of them is required. Further, the inherently coarticulated nature of gestures means that the sounds of speech cannot be divided into segments that correspond to phonological segments. Speech is a complex code and not a simple cipher. As Liberman and colleagues have often noted, in the uttering of *bag* "coarticulation folds three phonological segments into a single seamless stretch of sound in which the information about the several phonological segments is thoroughly overlapped" (A. M. Liberman, 1992, p. 175). In terms of this latter example, what the would-be reader must know explicitly in order to become competent in the alphabetic principle is that the single sound *bag* consists of three phonological components. At the

outset, this knowledge is available to the would-be reader only tacitly in the workings of the phonetic module.

It should be evident from this cursory summary of A. M. Liberman's (1996) views that further results of the kind reported in the present article would go far to establishing his central claim that theories of speech and reading are inseparable. The immediate value of Liberman's effort to ground the understanding of literacy in speech theory is that it rationalizes findings of an extraordinarily fast translation of print to phonology. On his account, acquiring the means to effect that fast translation would be the eminently sensible thing to do. Furthermore, for Liberman, the evidence that the translation occurs too fast for the production of an auditory image or visual percept (as in masked phonological priming experiments at very brief prime–target onset asynchronies) is as it should be. At its base, language by ear is not auditory, and, at its base, language by eye is not visual. Both language by ear and language by eye are based in gestures. For a thoroughgoing account of both, a gestural phonology is required.

We can conjecture, therefore, that the inner speech at Huey's (1908/1968, p. 122) "core of ordinary reading" is an abstract gestural dynamic. Whatever the theoretical worth of this conjecture, it should not detract from the fundamental finding of the present research: The visual processing of a word can be affected by the word's physical phonetic details. In modeling phonological involvement in visual word recognition, it will not suffice to implement rules that map letters, singly or in combination, to phonemic segments or to define whole-word representations in a phonological lexicon indifferent to physical phonetics (e.g., Coltheart et al., 2001). At a minimum, a phonetically informed phonology is implied by Abramson and Goldinger's (1997) pioneering work and by our corroboration of their principal experimental finding. The research challenge now becomes that of specifying the particulars of the phonetically informed phonology that mediates reading and determining the generality of reading's basis in that phonology. That generality, we note, is presumed by mastery of the alphabetic principle (I. Liberman et al., 1989).

References

Abramson, M., & Goldinger, S. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics, 59,* 1059–1068.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* (Release 2) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36(A),* 233–252.

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

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 39(A),* 467–478.

Boersma, P. (1998). *Functional phonology (LOT International Series 11).* The Hague, the Netherlands: Holland Academic Graphics.

Browman, C., & Goldstein, L. (1986). Toward an articulatory phonology. *Phonology Year Book, 3,* 219–252.

Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6,* 201–251.

Browman, C., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics, 18,* 299–320.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155–180.

Browman, C., & Goldstein, L. (1995). Dynamics and articulatory phonology. In R. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 175–194). Cambridge, MA: MIT Press.

Browman, C., & Goldstein, L. (2001). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlee, 5,* 25–34.

Byrd, D. (1996). A phase window framework for articulatory timing. *Phonology, 13,* 139–169.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica, 22,* 125–159.

Clements, G. N. (1992). Phonological primes: Features or gestures? *Phonetica, 49,* 181–193.

Clements, G. N., & Hume, E. V. (1995). The internal organization of speech sounds. In J. A. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 245–306). Cambridge, MA: Blackwell.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review, 100,* 589–608.

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.

Davelaar, E. E., Coltheart, M., Besner, D., & Jonasson, J. T. (1978). Phonological coding and lexical access. *Memory & Cognition, 6,* 391–402.

Egger, V. (1904). *La parole intérieure: Essai de psychologie descriptive. Thèse présenté à la Faculté des Lettres de Paris par Victor Egger* [Inner speech: An experiment of descriptive psychology. Thesis presented by Victor Egger to the Faculté des Lettres, Paris] (2nd ed.). Paris: Alcan.

Flemming, E. (1995). *Auditory representation in phonology.* Unpublished doctoral dissertation, University of California, Los Angeles.

Fowler, C. A. (1995). Speech production. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (pp. 29–61). New York: Academic Press.

Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory, 20,* 269–337.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–280.

Goldsmith, J. (1990). *Autosegmental and metrical phonology.* Oxford, England: Basil Blackwell.

Gotlob, L., Goldinger, S., Stone, G., & Van Orden, G. (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 561–574.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1331–1356.

Huey, E. B. (1968). *The psychology and pedagogy of reading.* Cambridge, MA: MIT Press. (Original work published 1908)

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretical approach.* Unpublished doctoral dissertation, University of California, Los Angeles.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology.* New York: Academic Press.

Kirchner, R. (2001). *An effort based approach to consonant lenition.* New York: Routledge.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America, 54,* 1102–1104.

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

Laeufer, C. (1992). Patterns of voicing-conditioned vowel duration in French and English. *Journal of Phonetics, 20,* 411–440.

Liberman, A. M. (1992). The relation of speech to reading and writing. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 167–178). Amsterdam: Elsevier Science.

Liberman, A. M. (1996). *Speech: A special code.* Cambridge, MA: MIT Press.

Liberman, A. M. (1998). When theories of speech meet the real world. *Journal of Psycholinguistic Research, 27,* 111–122.

Liberman, A. M. (1999). The reading researcher and the reading teacher need the right theory of speech. *Scientific Studies of Reading, 3,* 95–111.

Liberman, A. M., & Mattingly, I. G. (1989, January 27). A specialization for speech perception. *Science, 243,* 489–494.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (JARLD Research Monograph Series; pp. 2–33). Ann Arbor: University of Michigan Press.

Lukatela, G., Carello, C., & Turvey, M. T. (1990). Phonemic priming of words and pseudowords. *European Journal of Cognitive Psychology, 2,* 375–394.

Lukatela, G., Eaton, T., Lee, C., & Turvey, M. T. (2001). Does visual word identification involve a sub-phonemic level? *Cognition, 78,* B41–B52.

Lukatela, G., Frost, S. J., & Turvey, M. T. (1999). Identity priming in English is compromised by phonological ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 775–790.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: 2. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 331–353.

Lukatela, G., & Turvey, M. T. (1996). Inhibition of naming by rhyming primes. *Perception & Psychophysics, 58,* 823–835.

Lukatela, G., & Turvey, M. T. (1998). Reading in two alphabets. *American Psychologist, 53,* 1057–1072.

Mohanan, K. P. (1993). Fields of attraction. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 61–116). Chicago: University of Chicago Press.

Neely, J. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Perfetti, C. A., & Zhang, S. (1995). The universal word identification reflex. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 33, pp. 159–189). San Diego, CA: Academic Press.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227–248). Amsterdam: North-Holland.

Peterson, G., & Lehiste, I. (1960). Durations of syllabic nuclei in English. *Journal of the Acoustical Society of America, 32,* 693–703.

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

Port, R. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America, 69,* 262–274.

Rey, A., Ziegler, J. C., & Jacobs, A. M. (2000). Graphemes are perceptual reading units. *Cognition, 75,* B1–B12.

Saltzman, E., & Byrd, D. (2000). Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science, 19,* 499–526.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333–382.

Seidenberg, M. S. (1995). Visual word recognition: An overview. In P. Eimas & J. L. Miller (Eds.), *Handbook of perception and cognition: Speech, language and communication* (pp. 131–179). San Diego, CA: Academic Press.

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.

Steriade, D. (1997). *Phonetics and phonology: The case of laryngeal neutralization.* Unpublished manuscript, University of California, Los Angeles.

Stevens, K. N., & House, A. S. (1963). Perturbations of vowel articulation by consonantal context. *Journal of Speech and Hearing Research, 6,* 111–128.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123,* 34–80.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1269–1291.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory & Cognition, 13,* 557–572.

Appendix

A Consideration of Autosegmental Phonology

Whereas features in the linear phonologies are independent properties of the individual segment or phoneme, features in autosegmental phonology are held separate from the segments. The features are represented as independent *tiers* that are hierarchically organized according to a principle that higher level features participate jointly in fewer phonological processes than lower level features. Features on the same tier are sequentially ordered. Features on different tiers are unordered and related to each other by means of so-called association lines that establish patterns of alignment and overlap (Clements & Hume, 1995).

Among the reasons for concluding that a feature should be assigned to an organizational level (a horizontal tier) rather than to a phoneme are phonological phenomena such as, for example, pronouncing the suffix *ed*

as [t] in *sipped* but as [d] in *sobbed*. In this example, the voicing feature is spread from the final consonant of the stem to the affix; thus, /p/ is voiceless and, therefore, /d/ is voiceless. Vowel quality and tones provide similar examples of a spreading across constituent phonemes.

Conventionally, lax vowels /I/ and /ɛ/ are characterized as *short*, and tense vowels /i/ and /e/ are characterized as *long*. In the underlying form of /bId/, /I/ is short and can be represented in the autosegmental formalism as an association with a single "timing slot" (Clements & Hume, 1995; J. Goldsmith, personal communication, February 2003). In the surface form of /bId/, /I/ is a relatively long vowel that can be encoded by association lines to two skeletal positions (Clements & Hume, 1995; Goldsmith, 1990). The underlying and derived forms are depicted in Figure A1. Crucially missing from the process depicted in Figure A1 is a motivating factor for the lengthening of the underlying short vowel. The motivating factor is not an assimilative process like spreading (see above), as there is no feature [± long] linked to the coda consonant in the underlying form that can encode both vowel and consonant length. We could suppose that consonants that trigger the lengthening of an underlying *short* vowel have associated skeletal positions. An associated skeletal position would trigger lengthening of the vowel presumably because the position must be filled by some high-ranking constraint or unblocked rule. The resultant surface form of this process would be [bId], with what amounts to a contrastively long surface vowel as depicted in Figure A1.

Similar issues arise for the tense vowel /i/. Its autosegmental representation would involve associations with two skeletal positions. Applying the
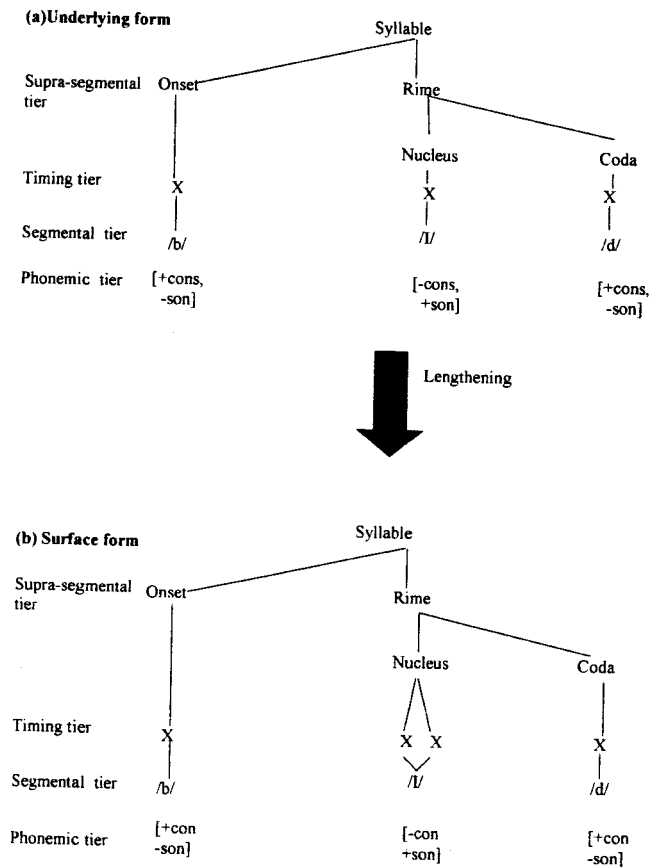

(a)Underlying form

Lengthening

(b) Surface form

Figure A1. Autosegmental description of (a) the underlying form and (b) the surface form of /bId/. The vowel lengthening effect for /I/ is captured by an expansion from one timing slot to two timing slots in the nucleus. *Cons* is an abbreviation for the feature consonant; *son* is an abbreviation for the feature sonorant; + and − signify, respectively, the presence and absence of the features.

(a) Underlying form

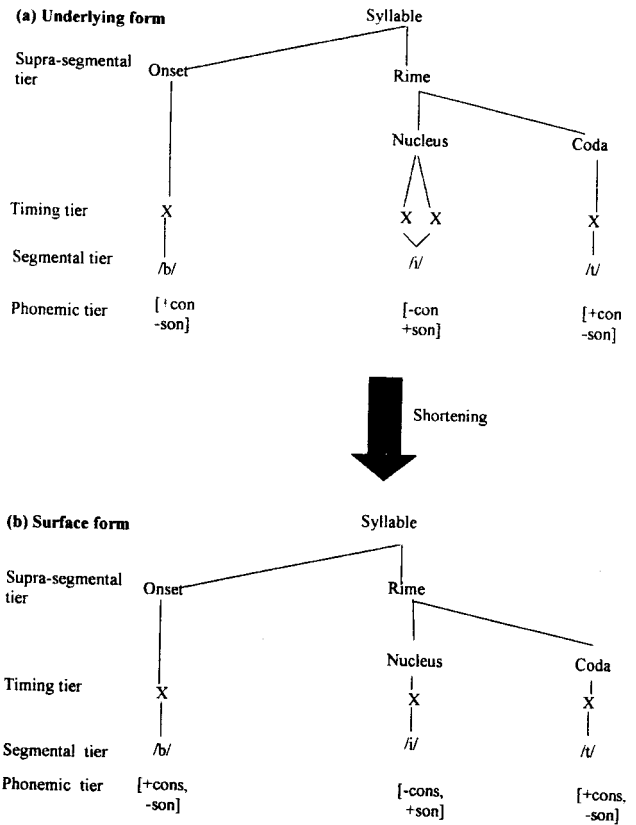Shortening

(b) Surface form

Figure A2. Autosegmental description of (a) the underlying form and (b) the surface form of /bit/. The vowel shortening effect for /i/ is captured by a reduction from two timing slots to one timing slot in the nucleus. *Cons* is an abbreviation for the feature consonant; *son* is an abbreviation for the feature sonorant; + and − signify, respectively, the presence and absence of the features.

above autosegmental account of vowel lengthening to /bid/, it could be presumed that, in the surface form, /i/ is associated with three skeletal positions. It is again a presumption that lacks a clear motivating factor.

A different option is to condense the rule or constraint in the following manner: $V \rightarrow \alpha$ long/___ α voice, obstruent (where the α before long and voice can represent [±] and here exist in identical states). In the case of /bid/, the rule applies vacuously, as the vowel is already long. Likewise in the case of /bIt/ the rule is satisfied by the underlying form. In contrast, for the cases of /bit/ and /bId/ a change from underlying to surface form would be required. Taking the tense vowel as an example, for /bid/ to be phonetically longer than /bit/ in the surface form, /bit/s underlying vowel must be shortened. This shortening could be implemented in the manner shown in Figure A2. However, by adopting the transformation depicted in Figure A2, we undercut the motivating argument made above for the vowel lengthening expressed in Figure A1. What we did in Figure A1 was ascribe the lengthening of the short lax vowel to a nonassociated (floating) segmental position introduced by a voiced coda obstruent. What we did in Figure A2 was ascribe the shortening of the long tense vowel to a segmental position associated with the vowel that was delinked due to the presence of a voiceless obstruent (in coda or in general). These two ascriptions are formally inconsistent.