Journal of Experimental Psychology:
Learning, Memory, and Cognition
2000, Vol. 26, No. 3, 719–750

0278-7393/00/$5.00 DOI: 10.1037//0278-7393.26.3.719

Strategic Control in Word Reading: Evidence From Speeded Responding in the Tempo-Naming Task

Christopher T. Kello and David C. Plaut
Carnegie Mellon University and Center for the Neural Basis of Cognition

To investigate strategic control over response initiation in word reading, the authors introduce the tempo-naming task. Relative to baseline performance in the standard-naming task, participants were induced to respond with faster latencies, shorter durations, and lower levels of accuracy by instructing them to time response initiation with an experimentally controlled tempo. The tempo response cue attenuated stimulus effects, and as faster tempos reduced latencies, the number of spelling-sound errors remained constant, whereas the number of word, nonword, and articulatory errors increased. To explain these results, the authors propose input gain as a mechanism of control over processing speed. The experimenters sketch how input gain could account for the current results as well as for the results from stimulus-blocking experiments testing the route emphasis and time criterion hypotheses of strategic control.

It takes roughly 400 to 600 ms for a skilled reader to begin the pronunciation of a single, clearly printed word. This ballpark range comes from a long history of speeded word-naming studies in which participants had been asked to pronounce a printed word "as quickly and accurately as possible" (or some instructions to that effect). The speeded word-naming task has been used to examine a wide variety of theoretical issues, including processes that map orthography to phonology (Glushko, 1979; Seidenberg, Waters, Barnes, & Tanenhaus, 1984); organization of the lexicon (Forster & Chambers, 1973; Frederiksen & Kroll, 1976); semantic, phonological, and orthographic priming (Forster & Davis, 1991; Tabossi & Laghi, 1992; Taraban & McClelland, 1987); sentence and discourse processes (Hess, Foss, & Carroll, 1995; Trueswell, Tanenhaus, & Kello, 1993); reading impairments (Patterson & Behrmann, 1997; Stanovich, Siegel, & Gottardo, 1997); and reading acquisition (Lemoine, Levy, & Hutchinson, 1993; Manis, 1985). In each of these areas of research, a primary source of data has come from latencies in naming tasks. Therefore, understanding the processes responsible for the initiation of a naming response is of general importance for interpreting naming data across research domains.

The standard mode of thinking about the initiation of a naming response is as follows: A representation of pronunciation is built up over time and is based on the results of processing at one or more other levels of representation (e.g., lexical, semantic, orthographic, and syntactic knowledge; Coltheart, Curtis, Atkins, & Haller, 1993; Kawamoto, 1988; Plaut, McClelland, Seidenberg, & Patterson, 1996). When the pronunciation is resolved according to some criterion of completeness, the response is initiated. Often the exact nature of the criterion is left unexplained, but a common assumption is that a response is initiated as soon as an entire pronunciation is completed by some criterion (but see Kawamoto, Kello, Jones, & Bame, 1998). For example, activation or saturation thresholds have been used (e.g., Coltheart et al., 1993; Plaut et al., 1996).

One reason why issues of response generation are often neglected is what Bock (1996) has termed the "mind in the mouth" assumption (p. 396). She argued that researchers often implicitly assume that articulation provides a relatively direct reflection of cognitive processing, but the link from cognition to behavior is mediated. For example, with regard to an activation threshold of pronunciation readiness, different participants, or even the same participants across trials, may set the threshold at different levels as a function of trading speed for accuracy (Colombo & Tabossi, 1992; Lupker, Taylor, & Pexman, 1997; Stanovich & Pachella, 1976; Strayer & Kramer, 1994; Treisman & Williams, 1991). The fact that naming instructions are usually ambiguous as to emphasis on speed versus accuracy increases the likelihood of variability in threshold placement.

Threshold variability is often not considered an issue, in part because the more central processes driving activation of a pronunciation are thought to be relatively unaffected by shifts in response generation thresholds. More recently, however, researchers have argued that gaining a better understanding of response generation in naming is important for interpreting naming data and developing theories of the

Christopher T. Kello, Department of Psychology, Carnegie Mellon University, and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania; David C. Plaut, Departments of Psychology and Computer Science, Carnegie Mellon University, and Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania.

This research was supported by National Institute of Mental Health Grants MH47566 and MH19102. We thank Stephen Lupker, Max Coltheart, Jay McClelland, Alan Kawamoto, one anonymous reviewer, and the Parallel Distributed Processing research group for their helpful comments and suggestions. We also thank Ben Payne for his help in collecting data for Experiment 3.

Christopher T. Kello is now at the House Ear Institute, Los Angeles, California.

Correspondence concerning this article should be addressed to Christopher T. Kello, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057. Electronic mail may be sent to ckello@hei.org.

underlying cognitive processes (Balota, Boland, & Shields, 1989; Jared, 1997; Kawamoto et al., 1998; Kello & Kawamoto, 1998; Lupker, Brown, & Colombo, 1997).

One theoretical issue that has been informed by research focused on response generation is that of strategic control over processing routes in generating a pronunciation from print (Jared, 1997; Lupker, Brown, et al., 1997; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Rastle & Coltheart, 1999). Many researchers have proposed that subjects can strategically emphasize or deemphasize one of two available processing routes based on task demands (the *route emphasis hypothesis;* Herdman, 1992; Herdman, LeFevre, & Greenham, 1996; Monsell et al., 1992; Paap & Noel, 1991; Plaut et al., 1996). The route emphasis hypothesis does not distinguish whether a given processing route is actually emphasized or deemphasized. Rather, the hypothesis describes any situation in which the processing of one or both routes is changed such that one is privileged over the other. For example, if the stimuli in a word-naming task consisted of nothing but irregular words (e.g., *sure, pint, rouse,* etc.),[1] it would behoove subjects to deemphasize the sublexical route because this route may provide incorrect information on the irregular spelling-sound correspondences.

Monsell et al. (1992) tested the route emphasis hypothesis by dividing stimuli in a word-naming task into pure and mixed blocks. The pure blocks contained either all pseudowords or all irregular words (of mixed frequency in Experiment 1 and separated by frequency in Experiment 2). The mixed blocks contained both pseudowords and irregular words. Monsell and his colleagues found that irregular words were generally named faster in pure versus mixed blocks, and they interpreted this as evidence that subjects de-emphasized the sublexical route in pure blocks of irregular words, presumably to reduce interference from sublexical processing. However, one puzzling aspect of their results was that the pure block latency advantage was only reliable for blocks of high-frequency (HF), but not low-frequency (LF), irregular words (for similar results, see Andrews, 1982; Frederiksen & Kroll, 1976).

Lupker, Brown, et al. (1997) and Jared (1997) revisited these findings and argued for an alternative to the route emphasis account. They first noted that if one defines *deemphasis* as slowed processing times (of the sublexical route in this case, and regardless of changes in variance), then LF irregular words should have an equal or greater advantage in the pure block compared with HF irregulars. This is because processing times to pseudowords must overlap more with LF compared with HF words provided that the mean of the sublexical route processing times is greater than that of the lexical route (as suggested by previous findings; e.g., words are named faster than nonwords; Forster & Chambers, 1973). By contrast, studies have revealed a greater pure block advantage for HF irregular words. Lupker, Brown, et al. (1997) reran the Monsell et al. (1992) blocking experiment (with minor variations), and they replicated the pure block advantage for HF irregulars. Moreover, they found a statistically reliable pure block disadvantage for the LF irregulars. Lupker, Brown, et al. (1997) ran a second experiment to provide a further test of the route emphasis account, in which all of the stimuli contained regular spelling-sound correspondences. In this case, the sublexical route should have remained active both in the pure and in the mixed blocks, and therefore no blocking effect should have been found. Once again they found a pure block advantage for HF words (now regular), but unlike their first experiment, they found a pure block advantage for the LF words as well. Jared (1997) found analogous results except that she compared blocks mixed with pseudowords versus blocks mixed with LF inconsistent words. In summary, the results from Jared (1997) and from Experiments 1 and 2 of Lupker, Brown, et al. (1997) were not predicted by the route emphasis hypothesis.

To explain their results, Lupker, Brown, et al. (1997) recategorized the stimuli as fast or slow on the basis of the mean latency for each stimulus type in the pure blocks. The pseudowords and LF irregulars were slow, and the HF regulars and irregulars were fast (LF regulars were in the middle). The pattern of results could then be described as follows: Whenever fast and slow stimuli were mixed, response latencies increased for the fast stimuli but decreased for the slow stimuli relative to when those stimuli were in pure blocks. This insight lead Lupker and his colleagues to propose that the blocking manipulation prompted subjects to adjust a *time criterion* to initiate naming responses. The general idea is that subjects can, to some degree, set a time deadline relative to stimulus onset (Ollman & Billington, 1972). If the pronunciation is not fully activated by that time (i.e., an activation threshold is also in place), then the response may be initiated on the basis of whatever representation of pronunciation is available at that time (also see Meyer, Osman, Irwin, & Kounios, 1988).

To maintain a certain level of accuracy while responding quickly, subjects adjust the time criterion on the basis of the difficulty of the stimuli presented during the experiment. A pure block of fast stimuli allows for a quicker criterion than does a pure block of slow stimuli. When fast and slow stimuli are mixed, subjects set a middling time criterion: Thus, fewer HF but more LF words are hurried. This hypothesis embodies a speed–accuracy tradeoff (Pachella & Pew, 1968; Wickelgren, 1977), so it predicts an increase in errors to slow stimuli in mixed blocks.[2] This is, in fact, what Lupker, Brown, et al. (1997) found.

Motivation for the Current Study

The study by Lupker, Brown, et al. (1997) raises two issues in the current context. First, in its simplest form, a time criterion means that a response is initiated at a particular point in time, neither before nor after that point (aside from random fluctuations). Of course, this cannot be

[1]Coltheart et al. (1993) defined an irregular word as one that has one or more irregular grapheme-to-phoneme correspondences (GPC) as determined by a set of correspondence rules (see also Venezky, 1970).

[2]The complementary prediction for fast stimuli (i.e., more errors in the pure block) could not be verified because performance was at ceiling for those stimuli in both the pure and the mixed blocks.

the case because such a criterion would predict no effects of stimulus processing on reaction times. One way to amend the time criterion hypothesis is to combine it with two activation criteria, a minimum and a maximum: The minimum must be reached to initiate pronunciation, a pronunciation is always initiated when the maximum is reached, and the time criterion operates between the two. This more complicated version of the time criterion hypothesis would seem, in principle, to account for the blocking results referred to previously. However, it would be necessary to specify how the activation criteria are set in order to draw any clear predictions from this hypothesis, and to our knowledge, this has not been addressed.

The second issue raised by the Lupker, Brown, et al. (1997) study is that the time criterion hypothesis could not, on its own, account for all of their findings. In particular, Lupker, Brown, et al. estimated LF irregular words and pseudowords to be of comparable speed (i.e., difficulty) on the basis of mean latencies to these stimuli in the pure block conditions. In this case, the time criterion hypothesis predicts no blocking effect when comparing pure blocks of each type with a mixed block of LF irregulars and pseudowords. However, the results from Experiment 1 showed a pure block disadvantage for LF irregulars and a pure block advantage for pseudowords. Lupker and his colleagues proposed an additional lexical checking strategy that subjects invoked only (but not always) when words were present in the stimulus block. The fact that Lupker and his colleagues needed to invoke an additional mechanism raises the question of whether a more parsimonious alternative to the time criterion hypothesis could be proposed (for similar issues in decision response tasks, see Ruthruff, 1996).

We believe that the time criterion hypothesis is worthy of investigation for two main reasons: (a) It provides a novel explanation for stimulus-blocking effects, but a more explicit mechanism needs to be proposed, and (b) it can potentially be used to address the time course of phonological processing in word reading. In light of these reasons, we set two goals for the current study: (a) to formulate a more explicit mechanism of control over response timing and (b) to formulate a hypothesis of how pressure for speed relates to the time course of processing. The first goal was set in the service of investigating the time criterion hypothesis, and the second goal was set to explore a specific prediction made by current models of word reading. In the next section, we step through the logic behind this prediction and provide some computational support for it. We then present three word-naming experiments examining control over the initiation of a speeded naming response. In the General Discussion, we propose a mechanism of control over processing that could account for the current findings, as well as for effects of stimulus blocking like those found by Lupker, Brown, et al. (1997).

Implication of a Time Criterion for Models of Word Reading

The hypothesis of a time criterion suggests that to the extent that an experimenter can manipulate the subject's time criterion, one could investigate the time course of processing in a fairly direct manner. If the shifting of a time criterion is one type of speed–accuracy tradeoff (Pachella & Pew, 1968), then setting it earlier in time should cause an increase in naming errors (as it did in Lupker, et al., 1997). If subjects could shift the criterion very early in time, a very high error rate should ensue. Speech errors have served as a primary source of evidence for developing and testing models of speech production (Dell, 1986; Dell & Reich, 1981; Levelt, 1989), and one could use the same approach toward the study of word reading. In the current context, fast error responses could serve as a window into the early time course of processing.

Current models of word reading have an explicit time course of processing from stimulus onset to response generation, but predictions concerning the trajectory of processing have only been tested indirectly. For example, Kawamoto and Kitzis (1991) showed that both the interactive-activation model of word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), as well as a distributed model of lexical memory (Kawamoto, 1988), make a specific prediction concerning the time course of phonological activation in word reading. When processing an irregular word such as *pint*, the models showed a strong influence of the regular, incorrect pronunciation /pınt/ (to rhythm with *mINT*) early in the time course of processing (i.e., a regularization error). As activation settled to a fixed state, the models showed that the correct phoneme usually quashed activation of the incorrect phoneme, but only later in processing. This general hypothesis was supported in a naming experiment that they conducted: The mean latency of 16 regularized responses to the word *pint* was 601 ms, whereas the mean latency of 46 correct, irregular pronunciations was 711 ms.

The prediction made by Kawamoto and Kitzis (1991) is also a more general property of most existing connectionist models of word reading. In particular, if distributed representations of orthography, phonology, and semantics all interact with each other, then orthography can generate a phonological code through two interacting but distinct pathways: a direct orthography-to-phonology mapping (i.e., the nonsemantic route) and an indirect orthography-to-semantics-to-phonology pathway (i.e., the semantic route). We refer to this general class of model as the *triangle framework* of word reading (Harm, 1998; Kawamoto, 1988; Orden, 1991; Plaut et al., 1996; Seidenberg & McClelland, 1989).[3] In this framework, the sublexical resonance (i.e., correlational structure) between orthography and phonology is stronger than that between phonology and semantics because there is more structure in the former's mapping (Orden, 1991; Van Orden & Goldinger, 1994). Therefore, for an item that

[3]The term triangle framework is used to refer to this class of connectionist word-reading models because they implement lexical processing in terms of interactions among distributed representations of orthography, semantics, and phonology, which are typically drawn at the corners of a triangle. The term is not intended to apply to other models (e.g., Coltheart et al., 1993) that, coincidentally, may also be depicted in the shape of a triangle.

contains one or more exceptional (i.e., rare) spelling-sound correspondences, the nonsemantic route may generate the more common correspondences early in processing. For these items, the semantic route helps to override the influence of sublexical knowledge, and this tends to occur later in processing. The model of word reading presented by Zorzi, Houghton, and Butterworth (1998) is also likely to make the same prediction given its clear distinction between assembled and retrieved phonologies.

To provide some evidence that early regularization errors are in fact a characteristic of current connectionist models of word reading, we examined the time course of phonological representations in the attractor model of word reading presented by Plaut et al. (1996). As a rough approximation, we applied a simple time criterion to the model by halting its processing at successively earlier points in time and categorizing each response into four possible categories: correct response, regularization error, word error, and miscellaneous error. The details of this simulation are reported in *Simulation Methods and Results.* The results were as expected: The total number of errors increased as processing was halted at earlier points in processing, which included an increase in the number of regularizations.

We can also consider the time course of phonology in the dual-route framework. Intuitively, one might expect that a dual-route implementation, such as the dual-route cascade (DRC) model (Coltheart et al., 1993), would predict no increase in the proportion of regularization errors during the early stages of processing. This runs counter to the triangle framework's prediction. The dual-route prediction seems to arise because the lexical-route processing times are hypothesized to be faster, on average, than are the rule-route processing times. Therefore, one would expect that word errors, but not regularization errors, would increase in proportion during the earlier cycles of processing. However, the rule route in the DRC model processes the input string from left to right over time, and the irregular grapheme of a test word is usually the second or third in position from left to right (in a monosyllabic English word, the vowel is most likely to be irregular; Berndt, Reggia, & Mitchum, 1987). If the rule route has enough time to output at least the first few phonemes before the lexical route can significantly influence the computation of phonology, then regularization errors may occur as often as, or more often than, word errors in the early cycles of processing.

Similar to our analysis of the Plaut et al. (1996) attractor model, we examined the time course of processing in the DRC model.[4] The details are reported in *Simulation Methods and Results,* but to summarize, the results were similar to those of the attractor model: The number of regularization errors, as well as other error types, increased as responses were taken at earlier points in processing.

Simulation Methods and Results

We ran simulations with both the Plaut et al. (1996) attractor model (Simulation 3 in that article) and a current version of the DRC model (M. Coltheart, personal communication, May 1998). Both simulations were run with the test stimuli from Experiment 2 of the current study. To examine the time course of phonology, baseline latencies for each model under normal processing conditions were determined. The models were then tested again with the stimuli from Experiment 2, but processing was halted at a number of different points in time prior to the baseline latency for each model. The phonological representations active at these halting points were categorized into one of four possible categories: correct, word error, regularization error, and miscellaneous error. A word error was a phonological output that corresponded to a word in the model's training corpus but was not the target. A regularization error was a phonological output that corresponded to the GPC rules as defined by Coltheart et al. (1993) but was not the target (these errors were only possible for the irregular stimuli). Miscellaneous errors included all other phonological outputs that did not reach a criterion of correctness (defined in the following paragraphs for each simulation). Miscellaneous errors were not separated into nonword and articulatory errors because in these models, this distinction can only be drawn by an arbitrary threshold. Therefore, an increase or decrease in miscellaneous errors can be assumed to correspond to an increase or decrease both in nonword and articulatory errors. The attractor model was used from Plaut et al. (1996) because it is one of the few published instantiations of the triangle framework that has an explicit time course (i.e., uses continuous time units).

The Plaut et al. (1996) Attractor Model

The mean latency under normal processing conditions for the stimuli from Experiment 2 was 1.85 units of time. The error breakdown for each halting time is shown in Table 1. All three error types (word, regularization, and miscellaneous) were found to increase as responses were generated at earlier points in time.

The DRC Model (Coltheart et al., 1993)

The mean number of processing cycles under normal processing conditions for the stimuli from Experiment 2 was 98. Under normal conditions, processing was complete when the activation of one or more phonemes in each position-specific pool crossed a threshold of 0.43. The model's responses were determined by taking the most active phoneme at each position (including null phonemes, if these were the most active). The error breakdown for each halting time is shown in Table 1. All three error types (word, regularization, and miscellaneous) were found to increase as responses were generated at earlier points in time.

Simulation Discussion

The two simulations produced comparable results: The number of regularization errors, as well as other error types, increased as responses were taken at earlier points in

[4]We thank Max Coltheart for making available the output of the DRC model over processing cycles for our stimuli.

Table 1
Error Counts for the Attractor Model and for the DRC Model Categorized by Error Type and by Halting Time

	Attractor model					DRC model				
Error type	0.8	0.7	0.6	0.5	Total	65.0	60.0	55.0	50.0	Total
Word	6	7	9	11	33	7	11	9	11	38
	(26.1)	(22.6)	(23.7)	(21.2)	(23.4)	(9.3)	(12.8)	(11.3)	(13.3)	(11.7)
Regular	10	14	18	19	61	10	13	15	17	55
	(43.5)	(45.2)	(47.4)	(36.5)	(43.1)	(13.3)	(15.1)	(15.5)	(15.0)	(14.7)
Miscellaneous	7	10	11	22	50	58	62	71	81	272
	(30.4)	(32.3)	(28.9)	(42.3)	(33.5)	(77.3)	(72.1)	(73.2)	(71.7)	(73.6)
Total	23	31	38	52	144	75	86	95	109	365

Note. DRC = dual-route cascade. Numbers in parentheses are column percentages.

processing. The main difference in error patterns between the two simulations was that the attractor model produced a large proportion of regularization errors overall, but the DRC model produced a large proportion of miscellaneous errors. Most of the miscellaneous errors in the DRC model were a failure to activate the rightmost phoneme(s) to criterion. We do not draw any conclusions based on this difference in the simulations because it depends on the setting of criteria, which could be changed in future simulations.

These simulations show that current models of word reading can make explicit predictions about the time course of processing. The experiments reported in the current study were meant to, in part, explore the time course of processing in word reading.

Experiment 1

Our initial research question was twofold. First, how precisely can subjects control their timing of response initiation? A demonstration of their ability to control response timing (or lack thereof) would be potentially useful in formulating a more specific mechanism of control over response timing than those given by Lupker, Brown, et al. (1997) and Jared (1997). Second, can subjects initiate their responses substantially faster than they do in the standard-naming task? If so, the errors could help formulate an account of control over response timing as it relates to the time course of processing in the word-reading system.

To address these questions, we developed a novel methodology called *tempo naming*. Prior to the presentation of a letter string, subjects are presented with a series of evenly spaced auditory beeps accompanied by the incremental removal of visual flankers on the computer screen. The letter string is presented on the final beep, and the task is to pronounce the letter string such that the response is initiated simultaneously with the subsequent beep (which is not actually presented). The rate of beep presentation (i.e., tempo) can be increased or decreased to require subjects to respond more quickly or more slowly. Tempo naming is similar to *deadline naming* (Colombo & Tabossi, 1992; Stanovich & Bauer, 1978), in which subjects are simply told to respond more quickly if a given response is slower than a preset deadline. A version of the deadline paradigm analogous to tempo naming would instruct subjects to go no faster than the deadline, as well as no slower. However, tempo naming is distinct in two important respects. First, tempo naming gives an explicit and precise cue (the beeps and visual flankers) for when to initiate each response. Second, subjects receive quantitative feedback on every trial indicating the amount (in hundredths of a second) and direction (fast or slow) that the response was off tempo. Subjects are instructed to adjust the timing of their responses such that their feedback is as close to zero as possible on every trial, even at the expense of accuracy.

If subjects have a mechanism akin to a time criterion at their disposal, then they should place it with a fixed relation to tempo (to the best of their ability). Studies in finger tapping have shown that behavior can be entrained to a rhythmic cue (Kurganskii, 1994; Mates, Radil, & Poppel, 1992), although there is significant error and variability within and across subjects (Yamada, 1995). One strategy that subjects could adopt to perform the tempo-naming task is to entrain an "internal metronome" to the tempo and then synchronize the hypothesized time criterion with the rate of the internal metronome. The way that subjects use the tempo is a research question in itself, and we address this question to some extent. However, the primary goal of creating the tempo-naming task was to examine the mechanism of control over response timing (independent of its relation to the perception and processing of tempo), as well as the time course of phonological processing.

With regard to a mechanism of control over response timing, one extreme hypothesis is that subjects can base response initiation exclusively on some cue other than the target stimulus (i.e., the tempo in the current study). We refer to this as the cue-driven hypothesis of control over response initiation. It might seem that the delayed-naming task is a good test of this hypothesis because it is cue driven ("do not initiate a response until the cue is presented"). Not surprisingly, researchers have found that stimulus effects are generally reduced in the delayed-naming task (e.g., Balota & Chumbley, 1985; McRae, Jared, & Seidenberg, 1990; Savage, Bradley, & Forster, 1990) and are eliminated altogether in some cases (McRae et al., 1990). The persistence of stimulus effects in some delayed-naming experiments might seem like evidence against the cue-driven hypothesis of control (i.e., responses were presumably driven by factors other than the cue). However, delayed naming is not a sufficient test because it is not a purely cue-driven task: Subjects have the freedom to respond anytime after the cue. The tempo-naming task is purely cue driven because subjects are instructed to initiate a response in time with the tempo, no sooner nor later. If subjects can obey the tempo absolutely, then stimulus factors should have no effect on response timing.

To test this extreme hypothesis, we chose stimuli that varied along dimensions known to affect latencies in the standard-naming task: printed frequency and spelling-sound consistency (Jared, McRae, & Seidenberg, 1990; Seidenberg et al., 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985; see the Stimuli section of Experiment 1 for details on our use of consistency). If the cue-driven hypothesis is correct, then we should find no effect of frequency or consistency on response latencies even if the tempo is set such that subjects are induced to respond as fast or faster than their average latency in a standard-naming task.

We manipulated spelling-sound consistency for a second purpose as well: to examine the influence of sublexical spelling-sound correspondences as a function of response timing. As explained and supported previously, both the triangle and the dual-route frameworks predict an increase in the number of regularization errors as processing is halted at earlier points in time. If the tempo-naming task does indeed tap into earlier points in processing, then these models predict an increase in the number of regularization errors to exception words. To investigate this issue, tempos were set to be as fast or faster than each subject's baseline naming latency as determined by an initial block of standard-naming trials. The only guide we had to determine how much faster than baseline subjects should be induced to respond was a study by Colombo and Tabossi (1992). Using a response deadline, they induced subjects to respond more than 60 ms faster than baseline without any significant increase in error rate. We wanted to induce errors, so we set the maximum tempo to induce responses considerably faster than 60 ms below baseline (150 ms maximum). We explored a range of tempos that were faster than baseline, because we did not know how well subjects could perform the task.

We needed to consider one auxiliary issue in creating the tempo-naming task. How do the acoustic properties of an initial phoneme affect subjects' ability to time a given response with the tempo? It has been known for some time that such properties will affect naming latencies as traditionally measured (Sherak, 1982; Sternberg, Wright, Knoll, & Monsell, 1980). Kawamoto et al. (1998) showed that even when problems with the voice key were alleviated, acoustic energy from responses with plosive stops as the initial phoneme (e.g., /b,d,g,p,t,k/) began much later (i.e., about 60 to 100 ms) than did comparable responses with nonplosive initial phonemes. Timing in the tempo-naming experiments was measured acoustically online and given as feedback, and subjects were instructed to respond with the best possible timing as measured by their feedback. If subjects' time response initiation was based only on articulatory commands, then the acoustic timing (i.e., feedback) would be consistently slow for plosive compared with nonplosive initial phonemes. Alternatively, subjects might be able to time their responses on the basis of the onset of acoustic energy. If this does not depend on the type of acoustic energy (i.e., periodic versus nonperiodic, as in voiced versus unvoiced phonemes), then the type of initial phoneme would have no effect on timing. As a third alternative, one might find differences based on the type of acoustic energy (elaborated on in the *Results* of Experiment 1). This issue was not central to our line of investigation, so we simply included a mix of initial phonemes in the test words and within blocks (but controlled for initial phonemes across levels of the independent variables). We mention it here because it is important in interpreting certain aspects of the findings in the current set of experiments.

Method

Subjects

A total of 33 subjects participated in the experiment as part of a requirement for an undergraduate psychology course. Subjects reported being native English speakers with normal or corrected vision.

Stimuli

The test stimuli in the tempo-naming task were composed of 52 high-frequency exception (HFE) words, 52 low-frequency exception (LFE) words, and 52 low-frequency consistent words (LFC). An additional 13 of each stimulus type were also chosen for the standard-naming portion. For each stimulus type, 13 of the 52 words chosen for tempo naming were also included in standard naming for a total of 26 words of each type in standard naming. All test words were monosyllabic. The factors controlled for were initial phoneme, number of letters, and number of phonemes (HFE words had higher summed positional bigram frequencies than did LFE and LFC words). Words were chosen in triplets (one of each type), and each triplet was matched on the control factors as closely as possible (LFE–LFC pairs were also matched on frequency, and HFE–LFE pairs were also matched on body consistency). The mean values for the independent and control factors as a function of stimulus type are given in Table 2, and the triplets are given in Appendix A. Printed frequency was estimated using the Kučera and Francis (1967) norms.

Table 2
Means of Control Variables for the Standard- and Tempo-Naming Portions of Experiment 1, Categorized by Stimulus Type

	Standard naming			Tempo naming		
Variable	HFE	LFE	LFC	HFE	LFE	LFC
Word frequency	999.0	4.2	2.8	617.3	8.2	6.2
Bigram frequency	9492	5589	4561	8072	5107	5023
No. of letters	4.2	4.5	4.2	4.3	4.6	4.3
No. of phonemes	3.4	3.7	3.7	3.3	3.6	3.6

Note. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

Spelling-sound consistency is a general concept that captures a statistical relationship between sublexical orthographic units and their corresponding pronunciations (Plaut et al., 1996): the distribution of different pronunciations for a given orthographic unit (which can be measured by token or type; Jared et al., 1990). The consistency of a particular pronunciation for a given orthographic unit increases as the number of alternative pronunciations decreases. To quantify the consistency of monosyllabic words, researchers usually consider a single orthographic unit (i.e., the *body*, defined as the vowel plus any final consonants) even though the concept applies to other units as well.

To create as large a pool of exception words as possible, we used the concept of consistency in its general form rather than its use as a label for body consistency. We categorized stimuli as consistent or exceptional as a function of the number of alternative pronunciations for all orthographic units greater than or equal to the grapheme, and less than the word, in size.

In particular, the pronunciation of a contiguous orthographic unit was exceptional if it comprised less than 50% of all of the position-specific pronunciation types (based on the positions onset, vowel, and coda) summed across all monosyllabic English words. For example, the *i*, the *in*, and the *int* in *pint* are all exceptional (e.g., compare with *tick*, *bin*, and *hint*). As another example, the *i* and *in*, but not the *ind*, in *kind* are exceptional (e.g., compare with *bind*, *find*, *mind*).

A word was defined as exceptional if it contained one or more exceptional orthographic units. A word was defined as consistent if all orthographic units mapped to their most common pronunciation. For example, *hook* is exceptional because the grapheme /oo/ usually maps to the long vowel /u/. By the same logic, *spook* is also exceptional because the orthographic body *ook* usually maps to the short vowel /ʊ/.

Consistency is different from GPC regularity in two important respects. First, the concept of irregularity is based on graphemes, whereas the concept of consistency applies to multiple levels of orthographic structure. Second, irregularity is based on discrete, all-or-none criteria (i.e., rules), whereas consistency is based on a continuous measure of the statistical distribution of pronunciations. Despite these differences, 85% of our exception words were also irregular by GPC rules. In addition, even though our definition of consistency is more inclusive than the body definition of consistency, 85% of our exception words were body exceptional. Irregularity and consistency have been the focus of research in other studies (e.g., Glushko, 1979; Jared et al., 1990), but their differences are not important for the issue of strategic control over response initiation.

The standard-naming blocks included 2 filler words at the beginning of each block, and the tempo-naming blocks included 156 fillers mixed throughout the four blocks; 10 were placed at the beginning of each block, and 29 were interspersed throughout each block. Filler words were mono- and bisyllabic and ranged in frequency and consistency. Standard naming consisted of one practice block and two test blocks, and tempo naming consisted of one practice block and four test blocks. Test stimuli were evenly mixed and balanced across blocks and were counterbalanced across subjects. The order of trials within blocks was randomized for each subject under the constraint that 2 fillers began each standard-naming block, and 10 fillers began each tempo-naming block. Standard naming always preceded tempo naming, and the practice blocks began each portion of the experiment. The order of test blocks was counterbalanced across subjects (in a Latin-square design for the tempo blocks). An equal portion of each stimulus type appeared in each test block within standard and tempo naming, and fillers were divided equally among test blocks as well. The standard-naming and the tempo-naming practice blocks consisted of 10 and 40 fillers, respectively.

Procedure

Subjects sat in front of a 17-in. (43.18-cm) monitor, approximately 2 ft (0.61 m) away, and wore an Audio-Technical (Stow, OH) headset cardioid microphone. The microphone was positioned approximately 1 in. (2.54 cm) away and 2 in. (5.08 cm) down from the subject's mouth, and it was plugged into a Soundblaster 16-bit sound card (Creative Technology, Ltd., Milpitas, CA). Subjects were given written instructions for the standard-naming task and were asked to read them silently. Following this, the experimenter summarized the instructions, and any questions were answered. Subjects were instructed that they would see words presented in isolation on the monitor (the width of each letter subtended approximately 1.2° of visual angle) and that their task was to pronounce each word out loud as quickly and as accurately as possible. The level of recording was calibrated for each subject. Following calibration, the subject ran through the practice trials with the experimenter present to make sure that the subject performed the task correctly. The practice block was followed by two continuous blocks of test trials.

Each standard-naming trial began with a *READY?* prompt centered on the monitor. The subject pressed the space bar to begin the trial, upon which the prompt was replaced with a fixation point. The fixation point remained for 500 ms and was replaced by a single word. The word remained on the screen until a vocal response was detected. All stimuli (for tempo naming as well) were presented in lowercase letters in a large, distinct font (similar in appearance to Times New Roman) to minimize letter confusions. The time from word onset to the beginning of the next trial was a fixed 1,500 ms (i.e., the screen was blank for any remaining time after the response was detected). This was necessary because each vocal response was digitized and stored on the hard drive using the Runword software package (Kello & Kawamoto, 1998).

Immediately after each subject completed the standard-naming portion of the experiment, the mean latency of all of the test trials was calculated (excluding any responses faster than 200 ms but including any errors that the subject may have made). Naming latency was calculated using an acoustic analysis algorithm described in Kello and Kawamoto (1998). In brief, the algorithm is sensitive to increases in amplitude (e.g., to detect voicing) as well as to frequency of acoustic energy (e.g., to detect frication). Each subject's mean latency in the standard-naming task was set as the baseline tempo for the upcoming tempo-naming blocks. The four test blocks for each subject were assigned four different tempos: baseline and 50, 100, and 150 ms faster than baseline (B-0, B-50, B-100, and B-150, respectively). As noted previously, the order of test blocks was counterbalanced in a Latin-square design. Stimuli were rotated across subjects such that each subject saw each test word once in the tempo-naming blocks (¼ of the test words appeared in standard naming as well), and each test word appeared in every tempo and in every block order across subjects.

After completing the standard-naming blocks, each subject was given written instructions for the tempo-naming task. After reading them silently, the instructions were summarized, and any questions were answered. A paraphrasing of the instructions is as follows (also see Figure 1):

> Each trial will begin with a prompt followed by the presentation of five pairs of visual flankers. Then, 5 beeps will be played successively in a steady rhythm, and the pairs of flankers will disappear one by one with each beep. Upon presentation of the fifth beep, a word will appear in between

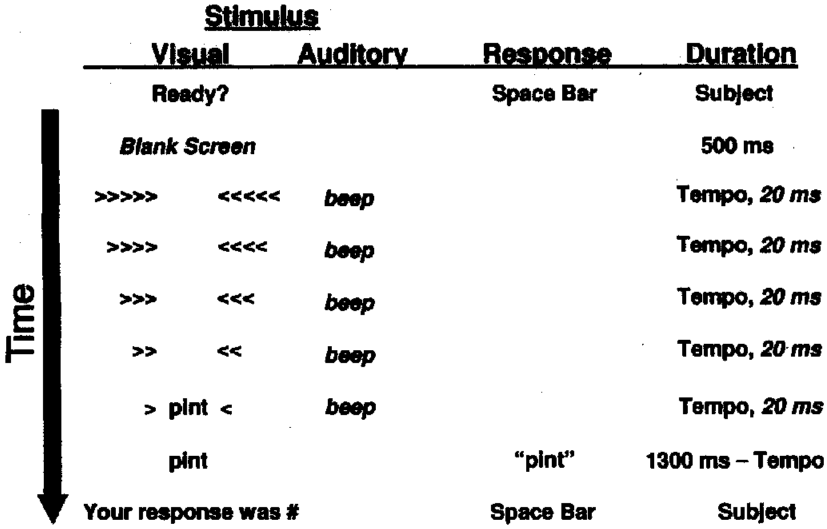

Figure 1. Diagram of the course of events for a single trial in the tempo-naming task. The "> <" symbols are flankers indicating the position of the target stimulus. Tempo is the time interval between each beep determined by the tempo condition and by the subject's baseline. "Subject" indicates that the duration is subject dependent.

> the last pair of flankers. Try to name the word such that the beginning of your response is timed with the sixth beep. However, no sixth beep will be played; your response should begin where the sixth beep would have been. You will get feedback after completing your response to tell you how well-timed it was to the tempo. The feedback is in the form of a number; the more positive it is, the slower your response; the more negative it is, the faster. A perfectly timed response produces a feedback of zero. Your primary task is to name the word on tempo, regardless of making errors. In the practice block you will see a mix of both relatively fast and slow tempos, but then you will run through four test blocks, and each one will be set at a different, but uniform, tempo.

After instruction, subjects ran through the practice block, which represented a randomly ordered but balanced distribution of the four tempos. The experimenter stayed with the subject through a number of trials to be sure the task was understood and to give any additional instruction if necessary. Each tempo-naming trial proceeded as follows: A *Ready?* prompt was presented in the center of the screen, and the subject pressed the space bar to begin. There was a 500-ms delay with a blank screen after pressing the space bar followed by the presentation of a paired set of flankers simultaneous with a brief tone that was 20 ms in duration. The flankers were sequentially erased from the outside inward in time intervals equal to the tempo for that trial. Each time a pair of flankers was erased, the tone was simultaneously presented. Interval durations were rounded up so that the removal of each flanker pair, as well as the presentation of each beep, could be synchronized with the video refresh rate (14-ms round up, at most). On presentation of the fifth beep (and removal of the fourth flanker pair), the target word was presented, centered between the last flanker pair. Recording from the microphone was initiated 200 ms before the next (sixth) interval and lasted for 1,500 ms. On the sixth interval, the last flanker pair was removed, and the word remained on the screen for the duration of recording. The word was then replaced with the message *Your response was #*, in which # equaled the amount of time in hundredths of a second that response latency differed from the sixth interval. This was computed by subtracting 200 ms from the calculated onset of acoustic energy relative to the onset of recording. The number was a positive or negative integer corresponding to the response offset from tempo. The feedback remained until the subject pressed the space bar, which brought up the *Ready?* prompt for the next trial. Subjects were explicitly asked to take a short break after the first two test blocks, and they were debriefed after completing all four test blocks.

Results

Standard Naming

Throughout the experimental results sections in this study, statistics over the standard-naming means and frequency counts (for error type analyses) are reported first, along with the magnitudes of any relevant effects. The tempo-naming results are reported afterwards, along with graphs including data from both the standard- and the tempo-naming results. The data are presented in this format to facilitate direct comparison of the standard- and tempo-naming means. Except for error type analyses, all statistics were analyses of variance (ANOVAs), and all analyses were conducted with subjects and items as the random factors (denoted as F_s and F_i, respectively, when presenting F values). Finally, all reported means in the current study were subject means unless stated otherwise.

Data removal. Data from 1 subject were removed because of equipment failure, and data from 1 item (*chic*) were removed because of an excess of errors (78%). Responses were removed if naming latency was less than 200 ms or greater than 1,200 ms. Responses were coded for errors (blind to the block that they appeared in), removed from all other analyses, and analyzed separately. In any cases in which multiple responses were given on a single trial, only the first one was considered for error categorization (but all such responses were considered errors of some type). Stutters that were followed by a fluent but incorrect response were categorized as incorrect responses rather than as articulatory errors. The error categories were as follows (examples were taken from the corpus of errors generated in Experiments 1 and 2).

1. *Word errors* were responses that formed a word but that did not match the target pronunciation. In all cases throughout this study, word errors were phonologically and/or orthographically similar to their targets (i.e., they differed in no more than two phonemes from the target). For example: *pint* → *pine, hitch* → *pitch,* and *glare* → *glad.*

2. *Legitimate alternative reading of components (LARC) errors* were responses to exception words that followed an alternate pronunciation of their exceptional orthographic unit and did not form a word. Strain, Patterson, Graham, and Hodges (1998) used the term in essentially the same way as we did, for example, *pint* to rhyme with *mint, mow* to rhyme with *now,* and *now* to rhyme with *mow.*

3. *Regularization and nonregularization errors* were two different types of LARC errors. Regularizations were those that followed GPC rules, and nonregularizations were the remainder. Dividing the LARC category in this way may be important for relating tempo-naming results to the DRC model of word reading.

4. *Mixed errors* were LARCs that also formed a word other than the target. For example, *great* → *greet, ghoul* → *goal,* and *plaid* → *played.*

5. *Nonword errors* were fluent pronunciations that did not form a word or a regularization. For example, *glove* → *guv, shoe* → *shope,* and *tunt* → *turt.*

6. *Articulatory errors* included all nonfluent pronunciations, in particular, stutters and garbled or incomprehensible responses.

Naming latency analyses. The main effect of stimulus type was significant by subjects and items, $F_s(2, 62) = 34.7$, $p < .001$, $F_i(2, 74) = 5.4$, $p < .01$. There was a nonsignificant, 7-ms decrease in mean latency from Block 1 to Block 2, $F_s(1, 31) = 1.1$, $p > .2$, $F_i(1, 74) = 2.9$, $p < .1$, and there was no reliable interaction of block and stimulus type, $F_s(2, 62) < 1$, $F_i(1, 49) < 1$. Pairwise comparisons revealed that a 26-ms advantage of HFE over LFE words (hereinafter referred to as a *frequency effect*) was significant by subjects and items, $F_s(1, 31) = 47.2$, $p < .001$, $F_i(1, 49) = 11.8$, $p < .001$. However, a 7-ms advantage of LFC over LFE words (hereinafter referred to as a *consistency effect*) was only reliable by subjects, $F_s(1, 31) = 5.1$, $p < .05$, $F_i(1, 49) = 1.3$, $p > .2$.

Error analyses. The main effect of stimulus type was significant, $F_s(2, 62) = 44.2$, $p < .001$, $F_i(2, 74) = 6.5$, $p < .01$, but there was no main effect of blocking, $F_s(2, 62) = 2.7$, $p > .1$, $F_i(1, 74) = 2.3$, $p > .1$. The interaction of block with stimulus was significant, $F_s(2, 62) = 6.0$, $p < .01$, $F_i(2, 74) = 5.4$, $p < .01$. Post hoc analyses showed that when collapsed across frequency, the error rate to exception words reliably decreased from Block 1 to Block 2, $F_s(1, 31) = 8.7$, $p < .01$, $F_i(1, 49) = 7.6$, $p < .01$, but marginally increased for regular consistent words, $F_s(1, 31) = 3.5$, $p < .05$, $F_i(1, 25) = 2.8$, $p < .1$. Planned comparisons showed that LFE words were reliably 9% more error prone than were LFC words (consistency effect), $F_s(1, 31) = 48.9$, $p < .001$, $F_i(1, 49) = 6.7$, $p < .05$, and 10% more error prone than were HFE words (frequency effect), $F_s(1, 31) = 59.5$, $p < .001$, $F_i(1, 49) = 7.5$, $p < .01$.

To provide more detail concerning errors, frequency counts were analyzed as a function of block and error type. Because the dependent measure is a frequency count, chi-square analyses were performed on the 2 × 5 contingency table formed by block and error type.[5] Collapsed across stimulus type, error counts were not reliably different than their expected values based on row and column means calculated across levels of block and error type, $\chi^2(4, N = 164) = 6.2$, $p > .15$.

Tempo Naming

Data removal. The subject removed from standard-naming analyses was also removed from tempo-naming analyses. Errors were coded in the same way as in standard naming (i.e., blind to block and therefore tempo) and were removed from all other analyses and treated separately (see *Error analyses*). Then, responses that were less than 175 ms or greater than 1,000 ms from the sixth tempo interval were removed (recording began 200 ms before the sixth tempo).

Latency analyses. Figure 2 graphs mean naming latencies (i.e., time from stimulus onset) as a function of stimulus type and tempo (including the mean latencies from the standard-naming task; note that the statistics presented here do not include standard-naming data; see Standard-Naming Results Section for those). The main effect of stimulus type was significant only by subjects, $F_s(2, 62) = 6.0$, $p < .05$, $F_i(2, 153) = 1.6$, $p > .2$, whereas the main effect of tempo was reliable in both analyses, $F_s(3, 93) = 214.3$, $p < .001$, $F_i(3, 459) = 245.0$, $p < .001$. The interaction did not reach significance, $F_s(6, 186) = 1.6$, $p > .1$, $F_i(6, 459) < 1$. Planned comparisons showed that the 5-ms frequency effect was reliable only by subjects, $F_s(1, 31) = 7.2$, $p < .05$, but not by items, $F_i(1, 102) = 2.3$, $p > .1$, whereas the 0.6-ms difference in latencies to LFE versus LFC words was not significant, $F_s(1, 31) < 1$ and $F_i(1, 102) < 1$. Planned comparisons of the tempo manipulation confirmed that each successively faster level of tempo caused responses to be reliably faster than the previous level: B-0 to B-50, $F_s(1, 31) = 130.1$, $p < .001$, $F_i(1, 153) = 104.9$, $p < .001$; B-50 to B-100, $F_s(1, 31) = 73.4$, $p < .001$, $F_i(1, 153) = 46.7$, $p < .001$; and B-100 to B-150, $F_s(1, 31) = 85.1$, $p < .001$, $F_i(1, 153) = 54.0$, $p < .001$.

These initial tests suggest that the influence of stimulus type on latencies is smaller in the tempo-naming task compared with the standard-naming task (a 26-ms frequency effect and a 7-ms consistency effect in standard naming vs. 5 ms and 0.6 ms in tempo naming, respectively). One possible reason for this attenuation of stimulus effects is that the overall variability in latencies decreased in tempo naming. This might be expected given that we asked subjects to respond at particular time intervals. However, inspection of

[5] In all of our chi-square analyses, the observations are not independent and may therefore be positively biased.

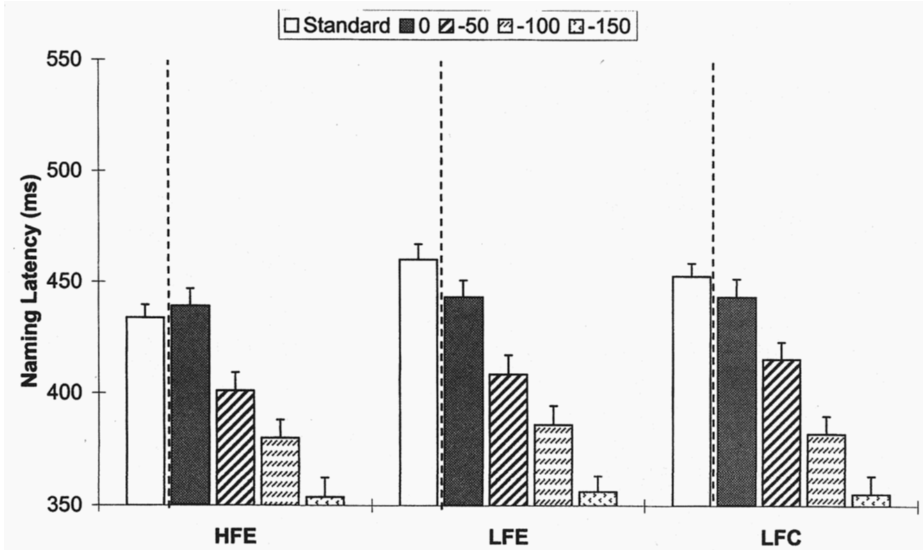

Figure 2. Mean latencies from the standard- and tempo-naming portions of Experiment 1 as a function of stimulus type and tempo. The dashed lines separate the standard-naming means from the tempo means. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

the standard error bars in Figure 2 shows that the within-cell variability was comparable across tasks (6.2 ms for standard naming, 7.9 ms for tempo naming, within-cell standard errors around the subject mean). An ANOVA with task as the independent variable and standard error as the dependent variable showed this difference to be nonsignificant, $F_s(1, 31) < 1$. Therefore, in terms of ANOVAs, the between-condition variance decreased in tempo naming but the within-condition variance did not.

However, there are three concerns with drawing the conclusion that between-condition variability decreased in tempo naming: (a) Standard naming always preceded tempo naming, (b) 25% of the tempo-naming stimuli also appeared in the standard-naming blocks, and (c) only half of the standard-naming stimuli appeared in tempo naming (the other half was not explicitly controlled against the tempo-naming stimuli). These three concerns were addressed as follows: (a) ANOVAs on tempo-naming latencies were conducted with the repeated stimuli removed, (b) the interaction of block order and stimulus type was examined to test for a practice effect within the tempo-naming task, and (c) the standard-naming latencies were reanalyzed with only those stimuli that appeared in tempo naming.

The tempo analyses with repeated stimuli removed were essentially identical to the analyses reported previously. Most relevant, the main effect of stimulus type was again reliable only by subjects, $F_s(2, 62) = 4.3$, $p < .05$, $F_i(2, 114) < 1$. The pairwise comparisons showed a 6-ms frequency effect that did not reach significance, $F_s(1, 31) = 2.4$, $p > .1$, $F_i(1, 76) = 1.1$, $p > .2$, and a 3-ms nonsignificant disadvantage for LFC words compared with LFE words, $F_s(1, 31) = 1.6, p > .2, F_i(1, 76) < 1$. If practice had reduced the effect of stimulus type on latency, then one would expect this effect to increase when repeated stimuli are removed; if anything, the effect decreased slightly (albeit, at least in part, because of reduced power). The analyses of block order and stimulus type revealed no discernible main effect of block order, $F_s(3, 93) < 1$, $F_i(3, 459) < 1$, nor Block Order × Stimulus Type interaction, $F_s(6, 186) < 1, F_i(6, 459) < 1$. If there was a practice effect from standard to tempo naming, one might expect this effect to continue through the blocks of tempo naming.

To illustrate the lack of a practice effect on latencies, Figure 3 shows the graph of naming latency as a function of block order and stimulus type. Finally, the reanalysis of standard latencies including only tempo-naming stimuli showed the same pattern of effects as the original analysis but with less power and therefore fewer significant comparisons. Relevant to the comparisons with tempo-naming results, the main effect of stimulus type was reliable by subjects, $F_s(2, 62) = 19.3$, $p < .001$, but marginally significant by items, $F_i(2, 36) = 2.7$, $p < .08$. Planned comparisons showed that the 23-ms frequency effect (cf. a 26-ms effect with all stimuli) was reliable, $F_s(1, 31) = 31.5$, $p < .001$, $F_i(1, 24) = 4.8, p < .05$, but the 5-ms consistency effect (cf. a 7-ms effect with all stimuli) was not, $F_s(1, 31) < 1$, $F_i(1, 24) < 1$. This final analysis suggests that the larger effect of stimulus type in standard versus tempo naming was not due to item differences.

Timing analyses. Naming latencies can also be graphed as offsets from perfect tempo. In other words, subjects were instructed to begin their response exactly on the sixth tempo interval, so one could graph their timing accuracy. ANOVAs on naming latency indicated that increases in tempo caused large, reliable decreases in naming latency. ANOVAs on timing would indicate whether response onsets were reliably different from tempo. Figure 4 shows the graph of mean timing offsets as a function of stimulus type and tempo. As

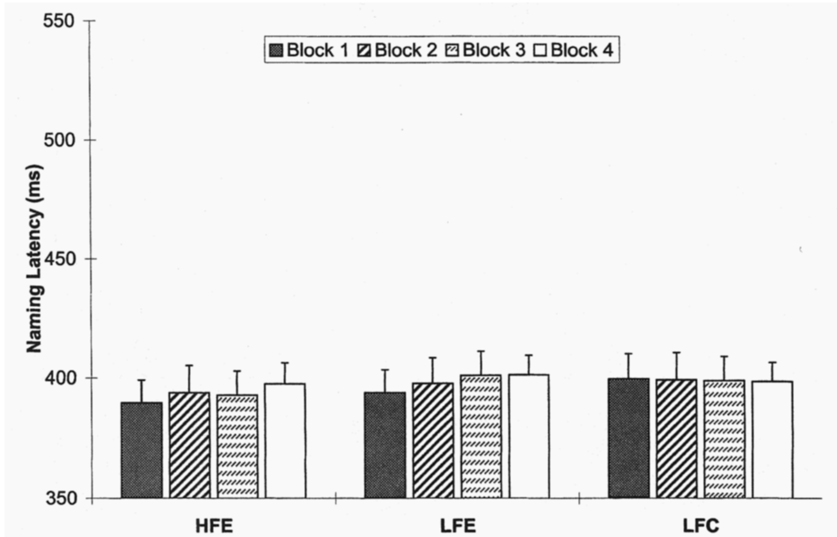

Figure 3. Mean latencies from the tempo-naming portion of Experiment 1 as a function of stimulus type and block. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

in the latency analyses, the main effect of stimulus type was only reliable by subjects,[6] $F_s(2, 62) = 6.0$, $p < .01$, $F_i(2, 153) = 1.3$, $p > .2$; the main effect of tempo was reliable, $F_s(3, 93) = 119.0$, $p < .001$, $F_i(3, 459) = 240.8$, $p < .001$; and the interaction was not significant, $F_s(6, 186) = 1.6$, $p > .1$, $F_i(6, 459) < 1$. Planned comparisons on stimulus type were not conducted because they are equivalent to the analogous comparisons with naming latency as the dependent measure. Planned comparisons to test whether timing was progressively delayed as tempo increased revealed that responses were in fact further delayed from tempo at each increase: B-0 to B-50, $F_s(1, 31) = 27.2$, $p < .001$, $F_i(1, 153) = 33.1$, $p < .001$; B-50 to B-100, $F_s(1, 31) = 70.0$, $p < .001$, $F_i(1, 153) = 70.7$, $p < .001$; B-100 to B-150, $F_s(1, 31) = 62.6$, $p < .001$, $F_i(1, 153) = 126.6$, $p < .001$. To characterize the failure of tempo to perfectly drive response initiation, a linear regression line was fit to the timing means at each level of tempo, averaged across stimulus type; the slope was 0.43 (i.e., 1 + *Srt*, in which *Srt* is the slope of the regression line for latency). If the tempo manipulation is perfect in determining response initiation, the slope would be 0.

One unexpected finding in the timing analyses was that responses were, on average, faster than tempo in the B-0 condition. If tempo conditions were mixed within blocks, this finding could have arisen from hysteresis of a response criterion (i.e., a relatively slow tempo trial preceded by a fast one might have a tendency to be overly fast; Lupker, Taylor, et al., 1997). However, because tempo was blocked and counterbalanced and each block began with 10 practice trials, this explanation cannot be correct. We reasoned that an adequate explanation would depend on the acoustic characteristics of response onset as well as on the answer to the question of what articulatory–acoustic marker subjects try to time with the tempo. We investigated this issue by examining timing as a function of tempo and initial phoneme type (graphed in Figure 5).

We categorized initial phoneme type on the basis of acoustic characteristics that are known to affect the measurement of response latencies (Kello & Kawamoto, 1998): voicing (voiced or unvoiced) and plosivity (plosive or nonplosive). Example words with an initial phoneme in each of the four categories are voiced plosives (*bed, deal, gate*), voiced nonplosives (*vet, zoo, red,* unvoiced plosives (*pet, tea, kite*), and unvoiced nonplosives (*fat, sea, thin*). Figure 5 shows that unvoiced initial phonemes, especially nonplosive ones, were fast in the B-0 condition, whereas voiced initial phonemes were closely timed to tempo in the B-0 condition. This result suggests that subjects, despite the numeric feedback, timed their responses with the onset of voicing (which we argue and statistically support in this article). To answer the question of why responses were faster than tempo in the B-0 condition, recall that the latency algorithm used in this study is sensitive to high amplitude and high-frequency acoustic energy. The onset of periodic energy is later in responses beginning with unvoiced compared with voiced initial phonemes relative to the onset of any type of acoustic energy. To accurately time the onset of voicing, measured latencies for unvoiced initial phonemes must be fast.

To argue for the hypothesis that subjects time responses with the onset of voicing, we shall distinguish it from the alternate hypotheses that timing was based on the onset of any acoustic energy (which was the basis of feedback), the

[6]In principle, significance levels for latency and timing analyses of stimulus effects (not of tempo or blocking effects) should be the same. However, they differ slightly because latency is relative to the onset of recording, whereas timing is relative to an estimate of the tempo interval. These estimates are not always aligned because of small variations in the onset of recording and small variations in the timing of the tempo interval due to alignment with the monitor refresh rate.

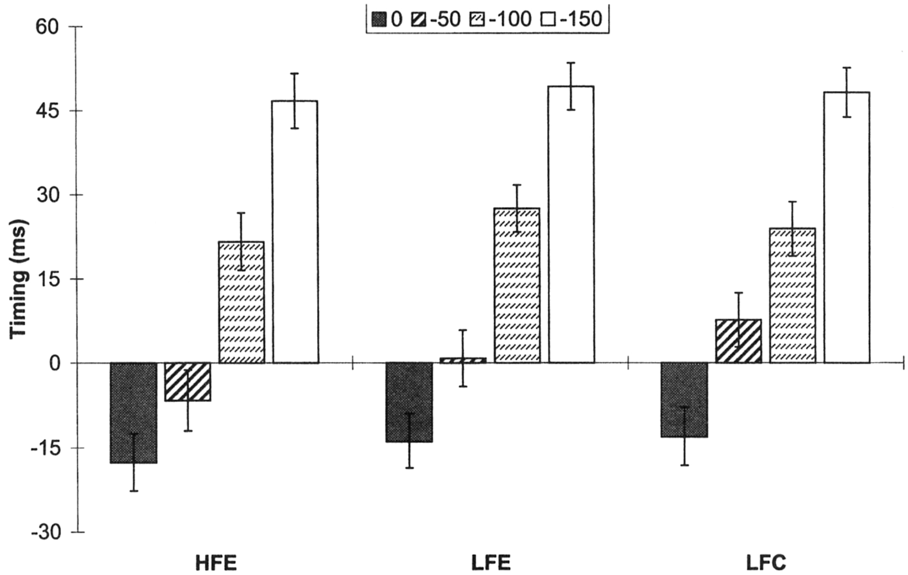

Figure 4. Mean timing accuracy from Experiment 1 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

onset of articulation, or the onset of the vowel. If timing was based on the onset of acoustic energy, there should be no effect of initial phoneme type. However, the timing of plosive initial words (i.e., /p,t,k,b,d,g,tʃ/;[7] $N_{item} = 63$) was 14 ms slower than that of nonplosive initial words ($N_{item} = 93$), $F_s(1, 31) = 33.2, p < .001, F_i(1, 154) = 16.5, p < .001$, and the interaction of plosivity and tempo was marginally significant by items, $F_s(3, 93) = 1.5, p > .2, F_i(3, 462) = 2.5, p < .06$. The effect of plosivity diminished slightly as tempo increased (18-ms effect at B-0, 17-ms effect at B-50, 10-ms effect at B-100, and 11-ms effect at B-150). The main effect of plosivity (as well as the other effects of initial phoneme reported later in this article) rules out the acoustic energy hypothesis, and we shall return to the interaction effect in the section on duration analyses. Next, if timing was based on the onset of articulation, then responses beginning with nonplosive phonemes should be relatively well timed (to the extent that subjects can keep up with the tempo), and those with plosive initial phonemes should be slow by comparison. This is because the onset of articulation more closely corresponds to the onset of acoustic energy for nonplosive-compared with plosive-initial responses (Kello & Kawamoto, 1998). However, plosive-initial responses were on tempo in the B-0 condition (timing = −2 ms), whereas nonplosive-initial responses were fast (timing = −18 ms), $F_s(1, 31) = 10.8, p < .05, F_i(1, 153) = 32.5, p < .001$. Finally, if timing was based on the onset of the vowel,[8] then there should be no effect of voicing of the initial phoneme because the vowel presumably begins at roughly the same point in comparable responses with voiced versus unvoiced initial phonemes. A post hoc split of the nonplosive-initial words by voicing on the initial phoneme revealed that responses to voiced stimuli were 23 ms slower than unvoiced stimuli, $F_s(1, 31) = 70.0, p < .001, F_i(1, 91) = 32.4, p < .001$. This effect of voicing interacted with tempo such that, as with plosivity, the effect diminished as tempo increased, $F_s(3, 93) = 2.4, p < .07, F_i(3, 273) = 6.6, p < .001$; we return to this effect in the following section. Furthermore, the mean timing of voiced, nonplosive-initial words in the B-0 condition was −3 ms, compared with −35 ms for comparable unvoiced words. The voicing onset hypothesis for the basis of timing in tempo naming explains both the voicing and the plosivity effects found, but the vowel onset hypothesis cannot explain the effect of voicing. Therefore, the data provide evidence for the voicing onset hypothesis.

Naming duration analyses. Researchers have shown that cognitive processes affect not only the onset of a naming response but its articulatory duration as well (Balota et al., 1989; Kawamoto, Kello, Higareda, & Vu, 1999; Kawamoto et al., 1998). For instance, Kawamoto et al. (1998, 1999) showed initial phoneme duration effects by contrasting the size of consistency effects in responses beginning with plosive versus nonplosive phonemes. They argued that a larger consistency effect on latency for plosive-versus nonplosive-initial stimuli (controlling for confounds) is evidence that consistency of the vowel affects the duration of the initial consonant(s). The logic was based on the premise (mentioned previously) that the acoustic onset of nonplosive-initial responses closely corresponds to the ac-

[7]The phoneme /tʃ/ (e.g., in the first phoneme in *chip*) is technically an affricate, but we treated it as a plosive because it has plosive-like characteristics.

[8]Previous research on the perceptual center of syllables (Hoequist, 1983; Marcus, 1981) suggests that vowel onset (or at least a correlate thereof) may play a role in synchronizing repeated syllables with a metronome.

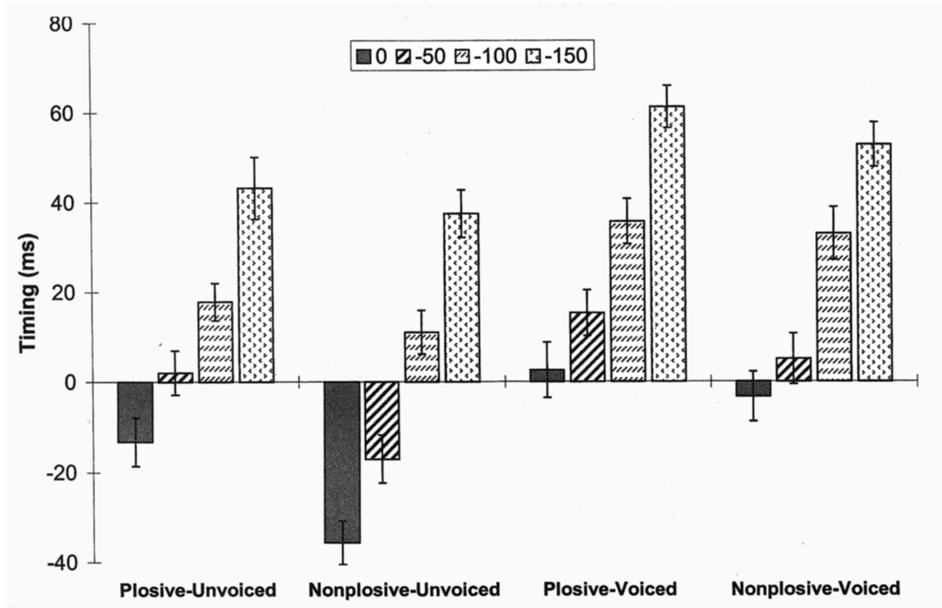

Figure 5. Mean timing with tempo from Experiment 1 as a function of tempo and the articulatory character of the initial phoneme (plosive vs. nonplosive and voiced versus unvoiced).

tual response onset, whereas the acoustic onset of plosive-initial responses conflates response onset with duration of the initial phoneme.

If we apply the same logic to the timing analyses conducted previously, then the interaction of stimulus type (with both voicing and plosivity) and tempo in the aforementioned timing analyses suggests that, at the least, naming durations for unvoiced, nonplosive-initial stimuli decreased as tempo increased. For example, if the duration of the initial phoneme caused the plosivity effect, then the weakening of this effect as a function of tempo indicates a decrease in initial phoneme duration. Similarly, if duration of the entire naming response was reduced for both plosive- and nonplosive-initial words, then the same interaction of tempo and plosivity would be predicted (analogous arguments could be made for voicing). We tested these alternate hypotheses by measuring whole-word naming durations (i.e., time from onset to offset of acoustic energy) as a function of initial phoneme voicing and plosivity, in conjunction with tempo.[9] The predictions were as follows. If tempo affected only initial phoneme duration, then there should be no effect of tempo on naming durations for plosive-initial responses. Because a duration effect on the initial phoneme in plosive-initial responses will mostly alter the silent gap in acoustic energy caused by pressure build-up for the plosive release, the duration effect will be reflected in naming latency rather than in acoustic duration. By contrast, if tempo affected whole-word durations, then all responses with any type of initial phoneme should show an effect of tempo.

The results unambiguously showed that tempo affected whole-word durations. Figure 6 shows the graph of naming durations as a function of tempo and initial phoneme type. Naming durations were calculated using the same algorithm for detecting the acoustic onset of a response, except that the algorithm was run backward from the end of response recording (Kello & Kawamoto, 1998; durations less than 50 ms or greater than 1,000 ms were removed from the analyses). Naming durations of plosive-initial responses showed a reliable effect of tempo, $F_s(3, 93) = 4.0, p < .01$, $F_i(3, 303) = 6.9, p < .001$. The main effect of tempo on naming durations, collapsing across initial-phoneme type, was also reliable, $F_s(3, 93) = 7.4, p < .001$, $F_i(3, 582) = 21.0, p < .001$. This effect did not significantly interact with voicing, $F_s(3, 93) < 1$, $F_i(3, 582) = 1.4, p > .2$, but it did interact marginally with plosivity, $F_s(3, 93) = 1.6, p > .2$, $F_i(3, 582) = 3.3, p < .05$. Qualitatively, naming durations of nonplosive-initial responses showed a stronger effect of tempo than did plosive-initial responses (which nonetheless showed a reliable effect of tempo). This interaction simply indicates that initial phoneme durations contributed to the effect of tempo on whole-word durations and that this contribution was attenuated for plosive-initial responses because plosive phonemes have much less acoustic extent than do nonplosive phonemes.

Error analyses. Errors were categorized as in the standard-naming analyses. Figure 7 shows the graph of mean error rate as a function of stimulus type and tempo. The main effect of stimulus type was reliable, $F_s(2, 62) = 38.2, p < .001$, $F_i(2, 153) = 9.4, p < .001$, as was the main effect of tempo, $F_s(2, 93) = 11.5, p < .001$, $F_i(3, 459) = 10.9, p <$

[9]The durations of most initial phonemes are difficult to measure (Kawamoto et al., 1998; Kello & Kawamoto, 1998), and we did not carefully choose stimuli with easily measured initial phoneme durations. We therefore chose whole-word durations as the dependent measure to examine.

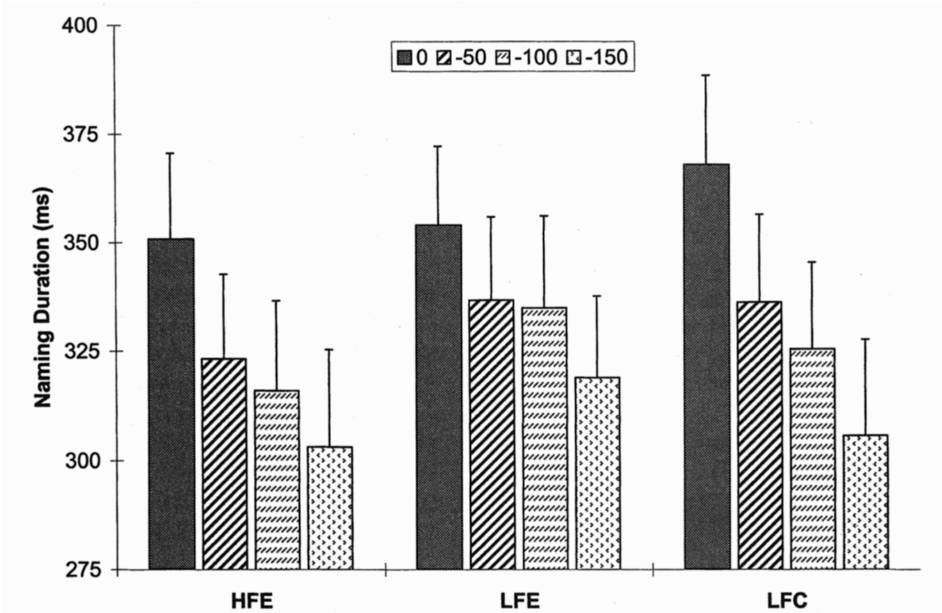

Figure 6. Mean acoustic naming durations from the tempo-naming task in Experiment 1 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

.001. As in the latency analysis, the interaction was again nonsignificant, $F_s(6, 186) = 1.5, p > .1, F_i(6, 459) = 1.4, p > .2$. Planned comparisons for stimulus type showed that LFE words were reliably more error prone than were HFE words, $F_s(1, 31) = 55.6, p < .001, F_i(1, 102) = 13.6, p < .001$, and likewise for LFE words compared with LFC words, $F_s(1, 31) = 43.6, p < .001, F_i(1, 102) = 9.4, p < .001$. Error rate results from planned comparisons over levels of tempo differed in part from those found in the latency analyses. Whereas each level of tempo was reliably faster than the previous level, only the increase from B-100 to B-150 showed a reliable increase in error rate: B-0 to B-50, $F_s(1, 31) < 1, F_i(1, 153) < 1$; B-50 to B-100, $F_s(1, 31) = 3.2, p > .05, F_i(1, 153) = 2.7, p > .1$; and B-100

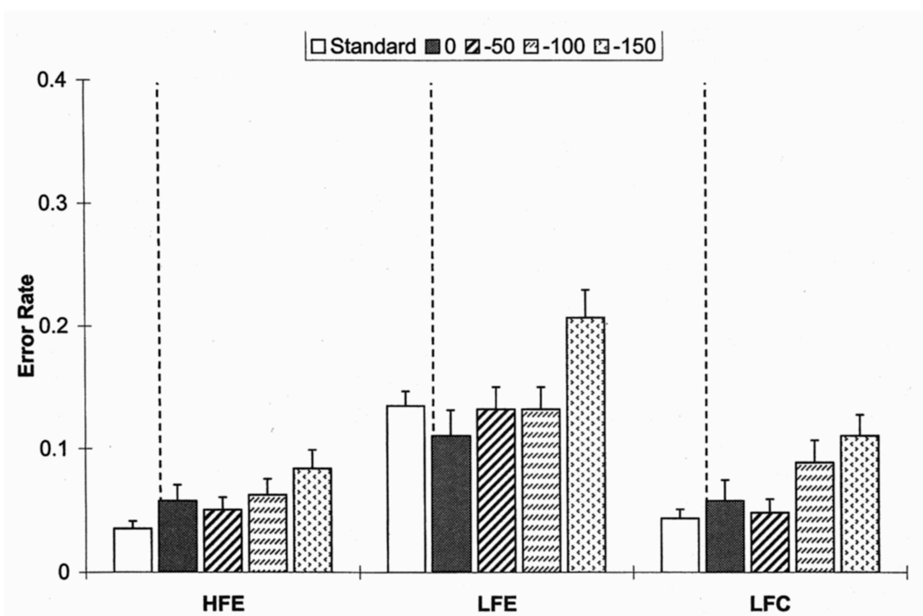

Figure 7. Mean error rates (proportion of errors) from the standard- and tempo-naming portions of Experiment 1 as a function of stimulus type and tempo. The dashed lines separate the standard-naming means from the tempo means. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

to B-150, $F_s(1, 31) = 10.2$, $p < .01$, $F_i(1, 153) = 8.2$, $p < .01$.

Errors were further analyzed by calculating chi-square statistics on a frequency table of tempo by error type, presented in Table 3. The overall frequency counts were significantly different than their expected values on the basis of row and column means, $\chi^2(12, N = 470) = 37.5$, $p < .001$. Our hypotheses concerned how the counts of different error types would vary as tempo increased. The Mantel-Haenszel (M-H) chi-square test for trend (Cody & Smith, 1997) is especially suited for contingency tables in which one or both variables have a specific ordering of levels (i.e., it collapses the independent contributions to degrees of freedom for individual levels of each variable). The M-H chi-square test for trend asks whether the cell counts in each of the N trend levels are increasing or decreasing in a uniformly linear fashion. The null hypothesis in the M-H test is that cell counts across trend levels are not changing linearly in proportion to each other. This is exactly the question we wanted to ask of our data: Is the proportion of LARC errors decreasing significantly as a function of tempo (the trend variable) relative to the number of other errors? The M-H test approached significance for the overall frequency counts, $\chi^2(1, N = ?) = 2.9470$, $p < .1$.

Motivation for analyzing error types specifically came from predictions concerning LARC errors. To address these more closely, we pooled LARC and mixed errors (which are LARCs themselves) together, and these were compared against the word errors: The M-H test was now significant, $\chi^2(1, N = 10.4551$, $p < .001$, as was the chi-square test, $\chi^2(9, N = 551) = 37.2$, $p < .001$. M-H analyses were also performed on the subset of data including only word, LARC, and mixed errors (with regularizations and mixed errors collapsed); the M-H test was significant, $\chi^2(1, N = 5.2551) = 5.2$, $p < .05$, but the chi-square test was not, $\chi^2(6, N = 551) = 5.8$, $p > .1$. On the basis of the pattern of column percentages in the frequency table, the chi-square results indicate that the frequency of LARC and mixed errors remained constant, whereas that of other error types increased as tempo increased. If we consider the mixed errors as mostly LARCs that coincidentally form words, then we can conclude that the proportion of LARC errors decreased as tempo increased.

We also divided LARC errors into true regularization errors (those following GPC rules) and nonregularization errors. Depending on how the tempo-naming task relates to the time course of processing, the DRC model and the attractor model of word reading may make different predictions concerning these error types. In particular, the DRC model predicts no nonregularization errors above chance (M. Coltheart, personal communication, May 1998). By contrast, the attractor model predicts mostly regularization errors and some nonregularization errors. The results showed that a substantial number of nonregularization errors did occur (25%), although the majority of LARC errors were regularization errors (75%). The proportion of regularization to nonregularization errors did not seem to change a function of tempo (the cell counts were too small to perform chi-square statistics). In order of increasing tempo, the regularization counts were 15, 14, 16, and 13, and the nonregularization counts were 5, 5, 4, and 5. The occurrence of nonregularization errors may be problematic for the DRC model, but this interpretation is dependent on how the tempo-naming task is operationalized in the model.

Discussion

The results of Experiment 1 can be summarized as follows. The manipulation of frequency and consistency in the standard-naming task basically replicated the findings of previous studies. Responses to HFE words were faster than those to LFE words, and responses to LFC words were faster than those to LFE words (reliable by subjects only). Error rates also showed this pattern, but more reliably. The tempo-naming task was effective in inducing progressively faster, more error prone responses (responses were 94 ms

Table 3
Frequency Counts of Errors in the Standard- and Tempo-Naming Tasks in Experiment 1, Categorized by Error Type and Tempo (for the Tempo-Naming Task)

		Tempo				
Error type	Standard	0	−50	−100	−150	Total
LARC	25	20 (22.7)	19 (19.8)	20 (16.7)	18 (10.8)	77 (16.4)
Word	38	33 (37.5)	43 (44.8)	54 (45.0)	60 (36.1)	190 (40.4)
Mixed	31	20 (22.7)	14 (14.6)	17 (14.2)	17 (10.2)	68 (14.5)
Nonword	31	10 (11.4)	8 (8.3)	21 (17.5)	39 (23.5)	78 (16.6)
Articulatory	39	5 (5.7)	12 (12.5)	8 (6.7)	32 (19.3)	57 (12.1)
Total	164	88	96	120	166	470

Note. Frequency counts were drawn from 2,496 test responses for standard naming (78 per subject) and 1,248 per tempo block (39 per subject per block). LARC = legitimate alternative reading of components. Numbers in parentheses are column percentages.

faster and 7% less accurate, on average, than baseline in the fastest tempo condition). This result indicates that, as Lupker, Brown, et al. (1997) noted, subjects must use a fairly conservative criterion (whatever the mechanism) to respond in the standard speeded naming task (otherwise one would expect tempo to affect speed less and accuracy more, relative to the observed effects).

The effect of stimulus type on naming latency diminished in the tempo-naming task compared with the standard-naming task, and this did not seem to be due to a practice effect (as indicated by blocking analyses) or to a change in items (as indicated by analyses of item subsets). Furthermore, there was no indication of an interaction of stimulus type with tempo in the latency analyses. An analysis of stimuli by plosivity and voicing of the initial phoneme indicated that subjects attempted to time the onset of voicing with the tempo. In addition, analyses of naming duration showed that as tempo increased, duration of the entire naming response decreased. Unlike naming latency, the pattern of error rate effects was essentially the same between standard and tempo naming: Error rates to HFE and LFC words were both lower than were those to LFE words. To complement the latency results, error rates increased with tempo, indicating a speed–accuracy trade-off (albeit the only reliable increase was from the B-100 to B-150 condition). An analysis of the error types as a function of tempo showed that whereas word errors and other error types increased in number with increases in tempo, the number of LARC errors remained constant.

The findings from Experiment 1 indicate the following in terms of the two main research agendas stated at the outset. First, subjects are quite good at timing the initiation of a naming response, as evidenced by the strong effect of tempo in even the fastest condition. However, the reduced but enduring frequency effect in the tempo-naming task challenges the strong cue-driven response hypothesis: Subjects were unable or unwilling to initiate a response on the basis of the cue alone. Note that the latency data are probably consistent with the time criterion hypothesis despite the evidence against the cue-driven hypothesis (depending on how one handles the accompanying activation criteria). Additionally, the latency data are consistent with a weak cue-driven hypothesis in which on some trials responses are based solely on the tempo, but on others responses are based solely on stimulus processing.

However, neither of these hypotheses account for the decrease in naming durations as tempo increased. These hypotheses simply do not address response execution, of which naming duration is a crude measure. One might argue that duration effects fall outside the scope of these hypotheses. However, understanding the factors underlying response duration sheds light on the processes underlying response initiation because they are, in fact, both integral parts of the generation of pronunciation. Therefore, we believe that an integrated account of effects on response initiation as well as an execution is desirable. In the General Discussion, we consider what type of mechanism might account for these data.

The second immediate research question was can subjects be driven to respond substantially faster than they did in the standard-naming task? The answer is clearly yes; in fact, because latencies reliably decreased by 27 ms from the B-100 to B-150 conditions, we suspect that subjects could be driven to respond correctly at even shorter latencies. The point of driving responses to be fast was to generate naming errors as a window into the time course of phonological processing. The proportion of LARC errors (which were mostly regularization errors) significantly decreased as tempo increased because of an increase in the occurrence of other error types (word errors most notably). If responses in the tempo-naming task reflected earlier stages of phonology in the normal course of processing, then the decrease in proportion of LARCs would seem to be problematic for both the triangle framework and the dual-route framework of word reading. As presented earlier, simulations with the Plaut et al. (1996) model and the DRC model (Coltheart et al., 1993) showed an increase in regularization errors as processing in the models was halted at successively earlier points in time. However, these results cannot be used as evidence for or against these models of word reading without specifying how the tempo-naming task affects processing. As a simplification, we used a strict time criterion in the simulations, but the persistence of a frequency effect on latencies indicates that a strict time criterion is incorrect. We return to this issue in the General Discussion.

Experiment 2

Experiment 2 was intended to replicate and extend the results of Experiment 1 and to test a route emphasis account of the error pattern in the tempo-naming portion of Experiment 1. The tempo-naming task is novel, so it is useful to know if the results from Experiment 1 can be replicated with an extended set of stimuli and a second group of subjects. More important, however, we need to explain the decrease in proportion of LARC errors with increased tempos. Thus far we have attributed this effect purely to the increase in pressure for speed. However, the effect could have arisen from strategic factors based on stimulus composition. Recall that one motivation for this study came from a debate concerning strategic effects in word naming. The route emphasis account proposed that subjects emphasize or de-emphasize one of the routes on the basis of composition of the stimulus list (Monsell et al., 1992), whereas the time criterion account proposed that subjects adjust a criterion to initiate pronunciation (Jared, 1997; Lupker, Brown, et al., 1997).

In Experiment 1, subjects may have de-emphasized the sublexical route because close to half of all of the stimuli in Experiment 1 contained exceptional spelling-sound correspondences, and processing in the spelling-sound route tends to interfere with the pronunciation of exception words. The fact that errors to LFE words decreased from Block 1 to Block 2 in the standard-naming task is consistent with the idea that subjects de-emphasized the spelling-sound route as they became familiar with the stimulus composition. To test this account, we included pseudowords as stimuli in Experiment 2. Following the logic laid out by Monsell et al. (1992), pseudowords should inhibit de-emphasis of the spelling-sound route because it is generally required for correct pseudoword performance. If the decrease in proportion of LARC errors across tempo in Experiment 1 was due to a strategic de-emphasis of the spelling-sound route, then the effect should be diminished when pseudowords are included. If, however, the rate of LARC errors continued to decrease with increased tempos, a route emphasis explanation would be discredited.

Method

Subjects

Thirty-four subjects participated in the experiment as part of a requirement for an undergraduate psychology course. Subjects reported being native English speakers with normal or corrected vision.

Stimuli

All test stimuli from Experiment 1 were included. In addition, 52 pseudowords were created by shuffling the onsets and bodies of the LFC words (listed in Appendix A). All 52 pseudowords appeared in the tempo-naming task, but none of these appeared in the standard-naming task; an additional 26 pseudowords were created to appear exclusively in the standard-naming blocks. Care was taken to avoid pseudohomophones (e.g., *brane*). Standard-naming and tempo-naming blocks of trials were created in the same way as in Experiment 1. An equal portion of each stimulus type appeared in each test block within standard and tempo naming, and fillers were divided equally among test blocks as well. There were 108 fillers in the tempo-naming task, and one fourth of these were pseudowords. The standard-naming and tempo-naming practice blocks consisted of 10 and 40 fillers, respectively, and one fourth of both practice blocks were pseudowords. Therefore, one fourth of all of the stimuli were pseudowords.

Procedure

The procedure was the same as that in Experiment 1 except that subjects were told that some letter strings were not legal English words. As with words, they were to name these letter strings as quickly and as accurately as possible.

Results

Standard Naming

Data removal. Data from 2 subjects were removed because of equipment failure, and data from 1 item (*chic*) was removed as in Experiment 1. Pseudoword errors were categorized as words were in Experiment 1 except that regularization and mixed were not possible (all pseudoword bodies contained consistent spelling-sound correspondences). Further data removal was carried out as in Experiment 1.

Naming latency analyses. There was a reliable main effect of stimulus type, $F_s(3, 93) = 25.8$, $p < .001$, $F_i(3, 99) = 11.1$, $p < .001$, but not of block, $F_s(1, 31) < 1$, $F_i(1, 99) < 1$. There was no reliable interaction of block and stimulus type, $F_s(3, 93) < 1$, $F_i(3, 99) < 1$. Pairwise comparisons revealed a reliable 18-ms frequency effect, $F_s(1, 31) = 17.3$, $p < .001$, $F_i(1, 49) = 7.2$, $p < .01$, and a reliable 15-ms consistency effect, $F_s(1, 31) = 15.1$, $p < .001$, $F_i(1, 49) = 5.1$, $p < .05$. The overall mean latencies of words compared with pseudowords was also tested (a lexicality effect); latencies to pseudowords were 35 ms slower than were latencies to words overall, and this difference was significant, $F_s(1, 31) = 36.0$, $p < .001$, $F_i(1, 101) = 24.7$, $p < .001$.

Error analyses. Errors were categorized and analyzed as in Experiment 1. There was a reliable main effect of stimulus type, $F_s(3, 93) = 31.7$, $p < .001$, $F_i(3, 99) = 10.4$, $p < .001$, but no main effect of block, $F_s(1, 31) < 1$, $F_i(1, 99) < 1$. The interaction of block with stimulus was significant by subjects, $F_s(3, 93) = 2.6$, $p < .05$, but not by items, $F_i(3, 99) = 1.7$, $p > .1$. Although the lack of a fully significant interaction prohibited post hoc analyses, the cell means clearly show a different pattern of results compared with Experiment 1: Whereas LFE errors decreased from Block 1 to Block 2 in Experiment 1, they remained constant in Experiment 2, and pseudoword errors decreased from Block 1 to Block 2. Planned comparisons showed a reliable 13% consistency effect on error rates, $F_s(1, 31) = 59.9$, $p < .001$, $F_i(1, 49) = 15.3$, $p < .001$, and a reliable 13% frequency effect, $F_s(1, 31) = 70.9$, $p < .001$, $F_i(1, 49) = 14.5$, $p < .001$. Finally, there was no reliable lexicality effect, $F_s(1, 31) < 1$, $F_i(1, 101) < 1$.

Frequency counts of error types by block were also analyzed. Cell counts were not significantly different than their expected values on the basis of row and column means calculated across levels of block and error type, $\chi^2(4, N = 135) = 4.7$, $p > .2$. To compare with Experiment 1, we also tallied counts with the pseudowords removed, and these too did not differ reliably from their expected values, $\chi^2(4, N = 135) = 3.4$, $p > .5$.

Tempo Naming

Data removal. The procedure for data removal was the same as that in Experiment 1.

Latency analyses. Figure 8 shows the graph of mean naming latencies as a function of stimulus type and tempo (including the standard-naming means). The main effect of stimulus type was significant, $F_s(3, 93) = 15.5$, $p < .001$, $F_i(3, 204) = 3.3$, $p < .05$, as was the main effect of tempo, $F_s(3, 93) = 439.2$, $p < .001$, $F_i(3, 612) = 246.0$, $p < .001$. The interaction did not reach significance, $F_s(9, 279) = 1.5$, $p > .1$, $F_i(9, 612) < 1$. Planned comparisons showed that the 9-ms main effect of frequency effect was reliable by subjects, $F_s(1, 31) = 16.3$, $p < .001$, but not by items, $F_i(1, 102) = 2.4$, $p > .1$, as was the 6-ms main effect of consistency effect, $F_s(1, 31) = 9.7$, $p < .01$, $F_i(1, 102) = 1.8$, $p > .1$. The 9-ms main effect of lexicality was reliable as well, $F_s(1, 31) = 24.6$, $p < .001$, $F_i(1, 206) = 7.0$, $p < .01$.

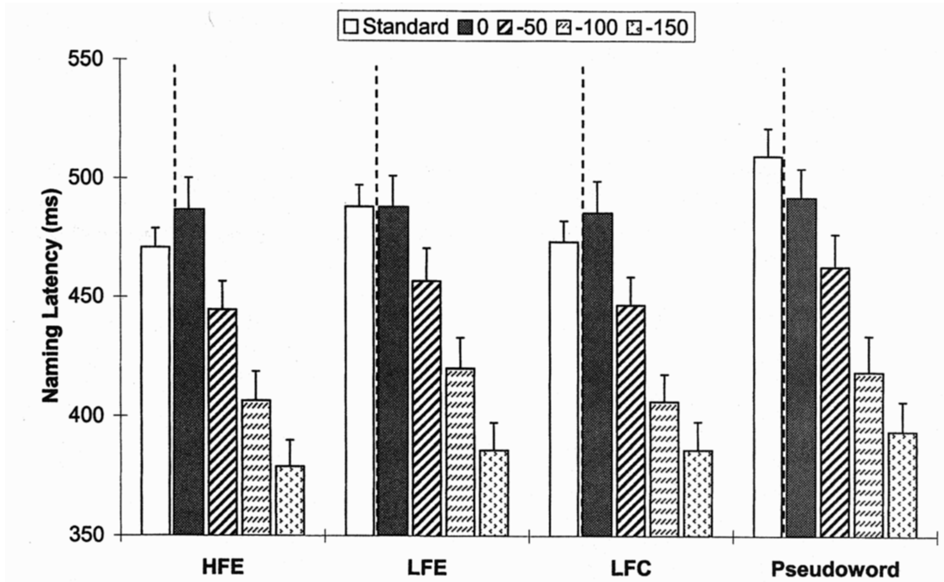

Figure 8. Mean latencies from the standard- and tempo-naming portions of Experiment 2 as a function of stimulus type and tempo. The dashed lines separate the standard-naming means from the tempo means. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

Planned comparisons of the tempo manipulation confirmed that each successively faster level of tempo caused responses to be reliably faster than the previous level: B-0 to B-50, $F_s(1, 31) = 223.8, p < .001, F_i(1, 207) = 94.5, p < .001$; B-50 to B-100, $F_s(1, 31) = 234.8, p < .001, F_i(1, 207) = 120.1, p < .001$; and B-100 to B-150, $F_s(1, 31) = 87.3, p < .001, F_i(1, 207) = 40.4, p < .001$.

As in Experiment 1, the influence of stimulus type on latencies was smaller in the tempo-naming task compared with the standard-naming task (an 18-ms frequency effect, 15-ms consistency effect, and 35-ms lexicality effect vs. 9-ms, 6-ms, and 9-ms effects in tempo naming, respectively). Also replicating Experiment 1 was the within-condition variability, which was not significantly different between the standard- and the tempo-naming tasks (9.5 ms vs. 12.6 ms, respectively, within-cell standard errors around the subject mean), $F_s(1, 31) < 1$.

The conclusion that stimulus effects were attenuated in tempo naming is supported by the following analyses (as in Experiment 1). First, we removed repeated stimuli (and all pseudowords because none of these were repeated) from the tempo-naming analyses, and the effects of stimulus type were essentially unchanged, albeit there was a loss of power by items (all effects were significant by subjects but not by items): the main effect of stimulus type, $F_s(2, 62) = 8.2, p < .01, F_i(2, 114) = 1.3, p > .2$; an 8-ms frequency effect, $F_s(1, 31) = 9.1, p < .01, F_i(1, 76) = 1.3, p > .2$; and a 9-ms consistency effect, $F_s(1, 31) = 12.4, p < .01, F_i(1, 76) = 2.2, p > .1$. Second, the analyses of block order and stimulus type once again revealed no discernible effect of block order, $F_s(3, 93) < 1, F_i(3, 612) = 2.1, p > .1$. Finally, we analyzed standard-naming latencies including only tempo-naming stimuli (thereby excluding all pseudowords), and the main effect of stimulus type was still reliable by subjects, $F_s(2, 62) = 5.7, p < .01$, but not by items, $F_i(2, 36) = 1.4, p > .2$. Planned comparisons showed that the 16-ms frequency effect (cf. an 18-ms effect with all stimuli included) was reliable by subjects only, $F_s(1, 31) = 8.8, p < .01, F_i(1, 24) = 2.9, p < .1$, as was the 12-ms consistency effect (cf. a 15-ms effect with all stimuli), $F_s(1, 31) = 7.5, p < .01, F_i(1, 24) = 1.7, p > .2$. Taken together, these three analyses indicate that the decreased stimulus effect on latencies in the tempo-naming task was due to the task itself rather than to practice or stimulus selection.

Timing analyses. Figure 9 shows the graph of mean timing offsets as a function of stimulus type and tempo. As in the latency analyses, the main effect of stimulus type was reliable, $F_s(3, 93) = 15.5, p < .001, F_i(3, 204) = 2.7, p < .05$, as was the main effect of tempo, $F_s(3, 93) = 86.2, p < .001, F_i(3, 612) = 165.7, p < .001$. The interaction of stimulus type and tempo was not significant, $F_s(9, 279) = 1.5, p > .1, F_i(9, 612) = 1.3, p > .2$. Planned comparisons showed that, as in Experiment 1, responses were increasingly delayed from tempo at each step: B-0 to B-50, $F_s(1, 31) = 28.1, p < .001, F_i(1, 207) = 32.2, p < .001$; B-50 to B-100, $F_s(1, 31) = 10.0, p < .01, F_i(1, 207) = 24.9, p < .001$; and B-100 to B-150, $F_s(1, 31) = 85.8, p < .001, F_i(1, 207) = 120.0, p < .001$.

The analyses of timing as a function of initial phoneme in Experiment 1 supported the hypothesis that subjects timed their responses on the basis of the onset of voicing. The main piece of evidence was that responses with voiced nonplosive-initial phonemes were timed more accurately than were those with unvoiced ones, which were too fast in the B-0 condition. Timing in Experiment 2 was also analyzed as a function of voicing and tempo: Responses to voiced nonplo-

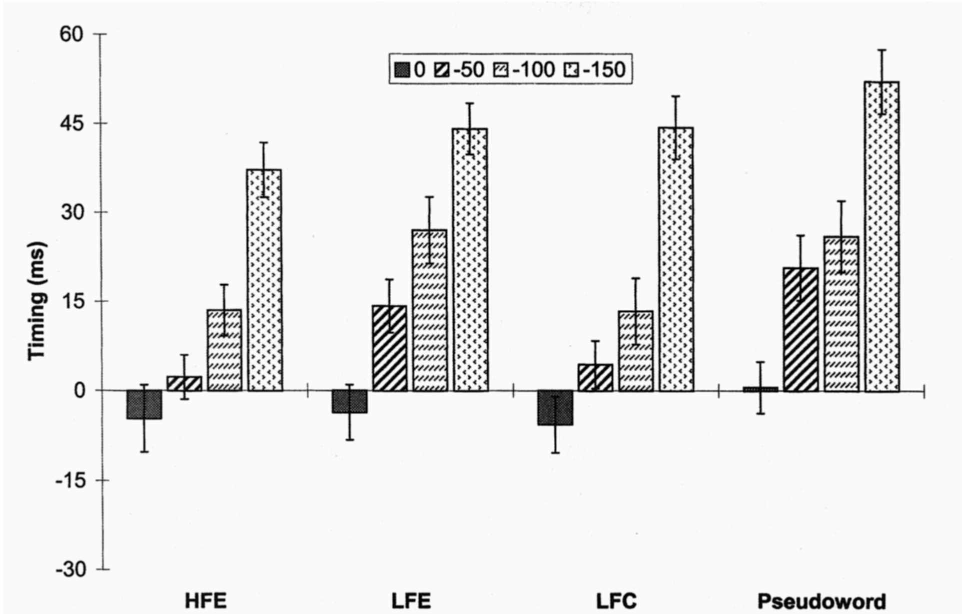

Figure 9. Mean timing with tempo from Experiment 2 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

sive-initial responses were 28 ms slower (cf. 23-ms effect in Experiment 1) than the unvoiced counterparts, $F_s(1, 31) = 104.3, p < .001, F_i(1, 123) = 54.1, p < .001$, and this effect interacted with tempo in the same way as in Experiment 1, $F_s(3, 93) = 3.3, p < .05, F_i(1, 123) = 3.5, p < .05$. In particular, the voicing effect steadily decreased from a 38-ms difference in the B-0 condition to an 18-ms difference in the B-150 condition. Finally, timing of nonplosive-initial voiced responses was 3 ms in the B-0 condition compared with −35 ms for the unvoiced responses. These analyses corroborate those from Experiment 1 in that subjects timed their responses on the basis of the onset of voicing rather than another candidates such as the onset of acoustic energy or the vowel. They also show indirect evidence of a decrease in naming duration as tempo increased.

Naming duration analyses. Naming duration analyses in Experiment 1 showed that an increase in tempo caused a decrease in whole-word naming duration for all initial phoneme types. Naming durations in Experiment 2 replicated the pattern found in Experiment 1 (shown in Figure 10: Durations, collapsed across initial phoneme type, steadily and reliably decreased from B-0 to B-150, $F_s(3, 93) = 10.3, p < .001, F_i(3, 612) = 20.6, p < .001$. The same pattern of effects held when analyses were restricted to voiced responses, $F_s(3, 93) = 4.0, p < .01, F_i(3, 177) = 6.7, p < .001$, as well as to plosive-initial responses, $F_s(3, 93) = 7.3, p < .001, F_i(3, 243) = 11.5, p < .001$. The fact that the naming duration effect held for both of the initial phoneme types mentioned previously indicates that the rime portion of the response decreased in duration as well as in the onset.

Error analyses. Figure 11 shows the graph of mean error rate as a function of stimulus type and tempo (including standard-naming means). The error rate results replicated those of Experiment 1. The main effect of stimulus type was reliable, $F_s(3, 93) = 55.5, p < .001, F_i(3, 204) = 14.9, p < .001$, as was the main effect of tempo, $F_s(3, 93) = 13.9, p < .001, F_i(3, 612) = 17.0, p < .001$. The interaction of stimulus type and tempo was nonsignificant, $F_s(9, 279) < 1, F_i(9, 612) = 1.3, p > .2$. Pairwise comparisons showed that the 7% increase in erred responses to LFE over HFE words was reliable, $F_s(1, 31) = 35.4, p < .001, F_i(1, 102) = 12.6, p < .01$, as was the 7% increase in LFE over HFE errors, $F_s(1, 31) = 60.4, p < .001, F_i(1, 102) = 15.5, p < .001$. To test the extent to which error rates increased with tempo as latencies decreased (i.e., a speed–accuracy trade-off), we compared error rates between adjacent levels of tempo. Although error rate increased with each increase in tempo, only the change from B-50 to B-100 was significant (cf. only the change from B-100 to B-150 was significant in Experiment 1): B-0 to B-50, $F_s(1, 31) = 1.9, p > .1, F_s(1, 207) = 2.3, p > .1$; B-50 to B-100, $F_s(1, 31) = 11.9, p < .01, F_i(1, 207) = 14.2, p < .001$; and B-100 to B-150, $F_s(1, 31) < 1, F_i(1, 207) < 1$.

Errors were further analyzed by calculating chi-square statistics on the frequency table of error counts broken down by tempo by error type (shown in Table 4). With LARCs and mixed errors treated separately, the overall frequency counts were not significantly different than their expected values in a chi-square test based on row and column means, $\chi^2(12, N = 853) = 17.8, p > .1$. However, when these two error types were combined, the chi-square test was significant, $\chi^2(9, N = 853) = 16.7, p < .05$. The purpose of analyzing error types here was to compare them with the analogous analyses in Experiment 1. In particular, we wanted to know whether the proportion of LARC errors fell as tempo increased (as they did in Experiment 1), so we conducted the

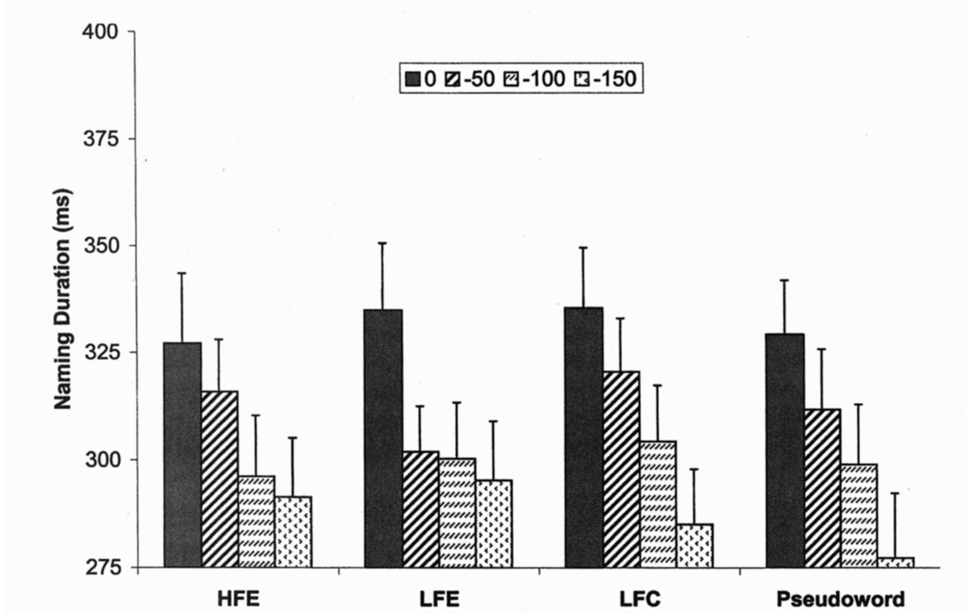

Figure 10. Mean naming durations from the tempo-naming portion of Experiment 2 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

M-H chi-square test. The M-H test was significant when comparing word errors, LARCs, and mixed errors for all stimuli, $\chi^2(1, N = 853) = 5.6, p < .05$ (LARCs and mixed errors were separated), $\chi^2(1, N = 853) = 8.3, p < .01$ (combined). The M-H was also significant when pseudowords were excluded from the analysis to compare with the corresponding results of Experiment 1, $\chi^2(1, N = 631) = 4.4, p < .05$ (separated), $\chi^2(1, N = 631) = 6.2, p < .05$ (combined). Finally, as in Experiment 1, most, but not all, LARC errors were regularizations. In addition, the proportion of regularizations did not change as a function of tempo. In order of increasing tempo, the regularization counts were 24, 23, 21, and 20, and the nonregularization counts were 4, 3, 10, and 5.

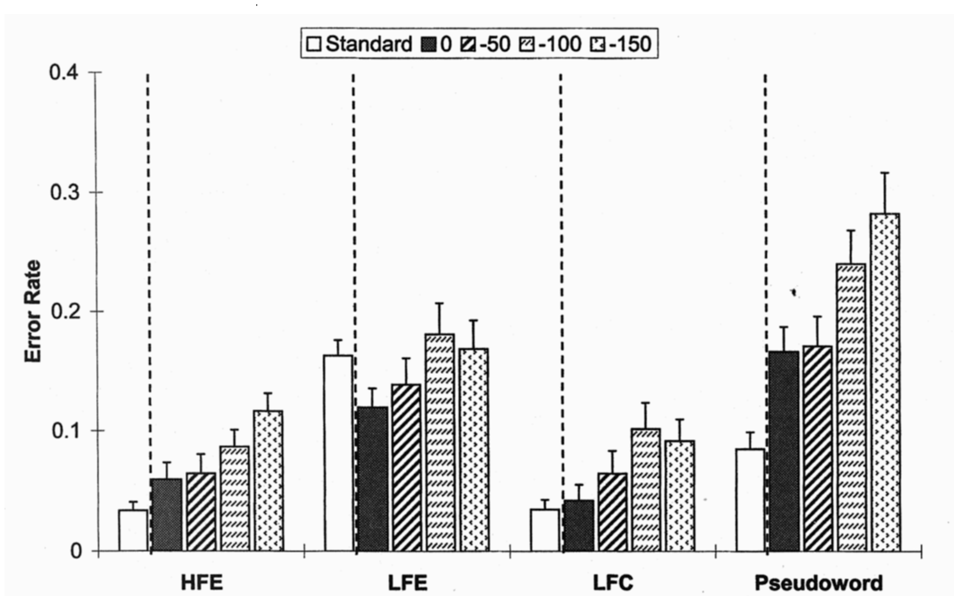

Figure 11. Mean error rates (proportion of errors) from the standard- and tempo-naming portions of Experiment 2 as a function of stimulus type and tempo. The dashed lines separate standard-naming means from the tempo means. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

Table 4
Frequency Counts of Errors in the Standard- and Tempo-Naming Tasks in Experiment 2, Categorized by Error Type and Tempo (for the Tempo-Naming Task)

Error type	Standard	Tempo 0	Tempo −50	Tempo −100	Tempo −150	Total
LARC	25	28 (17.7)	26 (14.9)	31 (12.8)	25 (9.0)	110 (12.9)
Word	31	63 (39.9)	70 (40.2)	108 (44.4)	128 (46.0)	369 (43.3)
Mixed	14	16 (10.1)	18 (10.3)	17 (7.0)	21 (7.6)	72 (8.4)
Nonword	48	39 (24.7)	39 (22.4)	54 (22.2)	57 (20.5)	189 (22.2)
Articulatory	17	12 (7.6)	21 (12.1)	33 (13.6)	47 (16.9)	113 (13.3)
Total	135	158	174	243	278	853

Note. Frequency counts were drawn from 3,328 test responses for standard naming (104 per subject) and 1,664 per tempo block (52 per subject per block). LARC = legitimate alternative reading of components. Numbers in parentheses are column percentages.

Discussion

Overall, the results of Experiment 2 replicated those of Experiment 1. Tempo again had a strong influence on response latencies and error rates (i.e., it induced subjects to incrementally trade accuracy for speed), but subjects did not fully keep pace with the tempo at the faster rates. Timing analyses showed again that subjects attempted to time their responses on the basis of the onset of voicing rather than on the onset of articulation or of the vowel. The frequency effect was diminished (but reliable) from standard to tempo naming, as was the consistency effect. Naming duration analyses again showed that durations decreased with increased tempo.

Two differences between the results of Experiments 1 and 2 were that (a) the consistency effect was more reliable in both the standard- and the tempo-naming portions of Experiment 2, and (b) the accuracy of naming LFE words dropped from Block 1 to Block 2 of standard naming in Experiment 2, whereas the accuracy of naming pseudowords increased; error proportions did not significantly change in the standard-naming blocks from Experiment 1. These results seem to indicate that as subjects saw more and more pseudowords during Block 1 of standard naming, they emphasized the sublexical route (and/or deemphasized the lexical route) to facilitate processing. However, the route emphasis hypothesis also predicts that LARC errors should increase as subjects trade speed for accuracy. To the contrary, the proportion of LARCs dropped with increases in tempo, replicating Experiment 1.

Taken together, the results from Experiments 1 and 2 provide more evidence against the route emphasis hypothesis as an explanation of the error patterns found in the tempo-naming task, but the results were mixed. In Experiment 3, we conducted a further test of the route emphasis explanation.

Experiment 3

If stimulus composition can cause subjects to preferentially emphasize the sublexical route (as was hinted at in the pattern of error rates in the standard-naming portion of Experiment 2), then perhaps the strongest way to encourage sublexical emphasis would be to present subjects with a stimulus block consisting of all nonhomophonic pseudowords. These items should rely almost exclusively on the sublexical processing route. The problem with using blocks of all pseudowords is that there would be no opportunity to observe LARC errors (i.e., for pseudowords with inconsistent bodies, all alternative pronunciations should be considered legitimate). The proportion of LARC errors served as a measure of emphasis placed on the sublexical route in Experiments 1 and 2. A stimulus block of all pseudowords can still be useful, however, because we can instead look for evidence of de-emphasis of the lexical route.

In particular, if word errors to pseudoword targets arise, at least in part, from processing in the lexical route, then such errors should decrease in proportion when the lexical route is de-emphasized compared with the proportion of word errors found in Experiments 1 and 2. Alternatively, if the lexical route is not de-emphasized in a block of all pseudowords, then the rate of word errors should be approximately equal to that found in Experiments 1 and 2. We tested these two possibilities in Experiment 3.

Method

Subjects

Thirty-six subjects participated in the experiment as part of a requirement for an undergraduate psychology course. Subjects reported being native English speakers with normal or corrected vision.

Stimuli

Out of the 52 pseudowords from Experiment 2 (all created by mixing LFC onsets and bodies), 48 were included in Experiment 3. The onsets and bodies of an additional 48 LFE words and 48 HFE words from Experiments 1 and 2 were mixed to create an additional 96 pseudowords for a total of 144 test pseudowords in Experiment 3 (all test pseudowords are given in Appendix B). An additional 50 pseudowords were created for the standard-naming task in order to measure a baseline naming latency for each subject. Standard-naming and tempo-naming blocks of trials were created in the same way as in Experiments 1 and 2. An equal portion of each pseudoword type (i.e., those created from LFC, LFE, or HFE words) appeared in each test block within tempo naming, and fillers were divided equally among test blocks as well. There were 40 pseudoword fillers in the tempo-naming task and 2 pseudoword fillers in the standard-naming task. The standard-naming and tempo-naming practice blocks consisted of 10 and 40 pseudoword fillers, respectively.

Procedure

The procedure was the same as that in Experiment 1 except that subjects were told that none of the letter strings would make legal English words. Subjects were instructed to name these letter strings as quickly and as accurately as possible.

Results

Standard Naming

Data removal. Data from 4 subjects were removed because of equipment failure. Further data removal was carried out as in Experiments 1 and 2.

Naming latency analyses. Mean latencies were analyzed as a function of block. A 15-ms increase in mean latency from Block 1 (538 ms) to Block 2 (553 ms) was not significant, $F_s(1, 31) = 1.8, p > .1, F_i(1, 49) < 1$.

Error analyses. Errors were categorized and analyzed as in Experiments 1 and 2. For pseudowords with inconsistent word bodies, all legitimate spelling-sound correspondences were coded as correct. A 2% increase in error rate from Block 1 (4%) to Block 2 (6%) was not significant, $F_s(1, 31) = 2.5, p > .1, F_i(1, 49) < 1$. Counts were not reliably different than their expected values based on means calculated across levels of block and error type, $\chi^2(2, N = 85) < 1, p = .2$.

Tempo Naming

Data removal. The procedure for data removal was the same as in Experiments 1 and 2 (there was no opportunity for LARC errors).

Latency analyses. Figure 12 shows the graph of mean naming latencies from the standard- and tempo-naming tasks as a function of tempo and stimulus type. Tempo once again had a strong influence on response latencies, $F_s(3, 93) = 345.1, p < .001, F_i(3, 423) = 166.9, p < .001$, but the main effect of stimulus type was significantly reduced compared with Experiments 1 and 2, $F_s(2, 62) = 3.4, p < .05, F_i(2, 141) < 1$. The interaction of stimulus type and tempo was not significant, $F_s(6, 186) < 1, F_i(6, 423) < 1$. Planned comparisons showed that the "pseudo-frequency" effect (i.e., latencies of pseudowords created from HFE vs. LFE words) was not reliable, $F_s(1, 31) = 2.8, p > .1, F_i(1, 94) < 1$, but the pseudo-consistency effect was reliable by subjects, $F_s(1, 31) = 6.4, p < .05, F_i(1, 94) < 1$. Planned comparisons of the tempo manipulation confirmed that each successively faster level of tempo caused re-

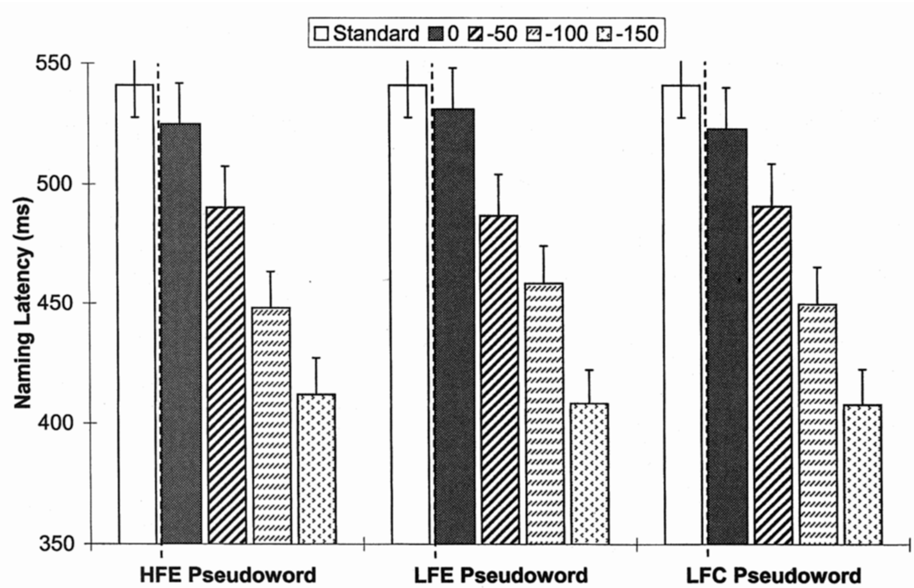

Figure 12. Mean latencies from the standard- and tempo-naming portions of Experiment 3 as a function of stimulus type and tempo. The dashed lines separate standard-naming means from the tempo means. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

sponses to be reliably faster than the previous level: B-0 to B-50, $F_s(1, 31) = 187.6, p < .001, F_i(1, 141) = 61.3, p < .001$; B-50 to B-100, $F_s(1, 31) = 128.4, p < .001, F_i(1, 141) = 46.5, p < .001$; and B-100 to B-150, $F_s(1, 31) = 150.8, p < .001, F_i(1, 141) = 71.0, p < .001$. The small, partially reliable pseudo-consistency effect was presumably due to the ambiguous spelling-sound correspondences that these strings contained (e.g., *bost* can rhyme with *cost* or *most;* Glushko, 1979; Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994; Taraban & McClelland, 1987).

Timing analyses. Figure 13 shows the graph of mean timing offsets as a function of stimulus type and tempo. As in the latency analyses, the main effect of stimulus type was only reliable by subjects, $F_s(3, 62) = 3.4, p < .05, F_i(2, 141) < 1$. In addition, the main effect of tempo was significant by subjects and by items, $F_s(3, 93) = 26.7, p < .001, F_i(3, 423) = 53.2, p < .001$, and the interaction was not significant, $F_s(6, 186) < 1, F_i(6, 423) < 1$. Planned comparisons showed that, as in Experiments 1 and 2, responses were increasingly delayed from tempo at each step: B-0 to B-50, $F_s(1, 31) = 18.7, p < .001, F_i(1, 141) = 25.8, p < .001$; B-50 to B-100, $F_s(1, 31) = 16.0, p < .01, F_i(1, 141) = 20.2, p < .001$; and B-100 to B-150, $F_s(1, 31) = 2.9, p > .1, F_i(1, 141) = 5.9, p < .05$.

Timing analyses were again conducted as a function of initial phoneme characteristics to test the hypothesis that subjects attempt to time their responses on the basis of the onset of voicing. As in Experiments 1 and 2, responses to voiced, nonplosive stimuli were reliably slower (32-ms difference) than were those to unvoiced, nonplosive stimuli, $F_s(1, 31) = 68.0, p < .001, F_i(1, 85) = 44.1, p < .001$, and this effect interacted with tempo, albeit reliable only by subjects, $F_s(3, 93) = 3.3, p < .05, F_i(3, 255) = 2.0, p > .1$. In particular, the voiced effect steadily decreased from a 40-ms difference in the B-0 condition to a 24-ms difference in the B-150 condition. Finally, timing of voiced, nonplosive items was 0 ms in the B-0 condition compared with −40 ms for unvoiced, nonplosive responses.

Naming duration analyses. Naming duration analyses were again conducted to show that the entire response decreased in duration as tempo increased rather than just the initial phoneme. The results replicated those of Experiments 1 and 2 (shown in Figure 14). Naming durations, collapsing across initial phoneme type, steadily decreased from B-0 to B-150, $F_s(3, 93) = 5.3, p < .01, F_i(3, 423) = 16.8, p < .001$. The same pattern of effects held when analyses were restricted to voiced, nonplosive pseudowords, $F_s(3, 93) = 2.8, p < .05, F_i(3, 117) = 6.3, p < .001$, as well as to plosive-initial pseudowords, $F_s(3, 93) = 4.0, p < .01, F_i(3, 165) = 5.6, p < .01$. The fact that the whole response duration effect held for both of the initial phoneme types mentioned previously indicates that the rime portion of the response decreased in duration as well as in the onset.

Error analyses. Figure 15 shows the graph of mean error rates for the standard and tempo portions of Experiment 3 as a function of stimulus type and tempo. Error rates again increased as tempo increased, indicating a speed–accuracy trade-off, $F_s(3, 93) = 5.4, p < .01, F_i(3, 423) = 8.8, p < .001$. However, unlike naming latencies, error rates were only marginally affected by stimulus type, $F_s(2, 62) = 2.7, p < .1, F_i(2, 141) < 1$. The interaction of stimulus type and tempo was also not significant, $F_s(6, 186) = 1.3, p > .2, F_i(6, 423) = 1.4, p > .2$. Error rates generally increased with each increase in tempo, but not every increase was reliable: B-0 to B-50 was not significant, $F_s(1, 31) < 1, F_i(1, 141) <$

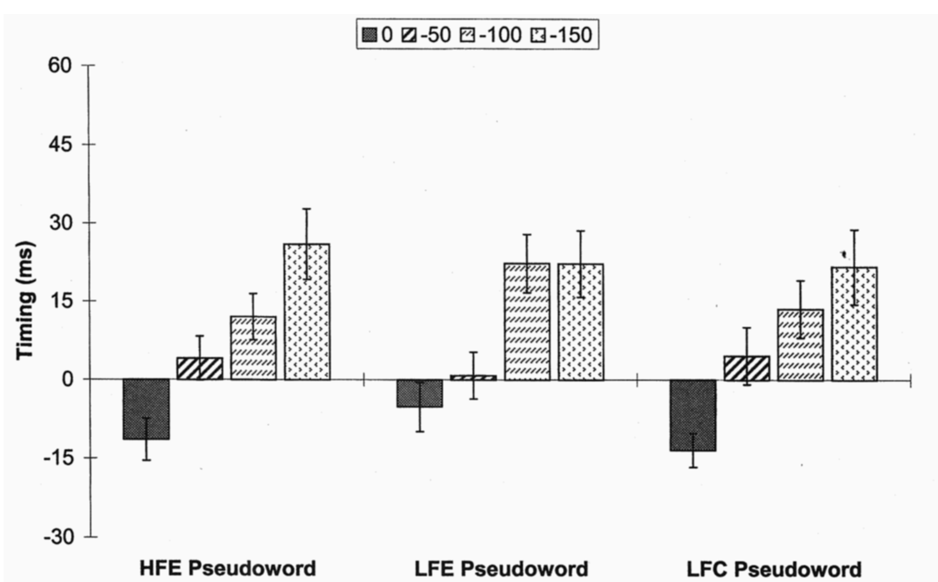

Figure 13. Mean timings with tempo from Experiment 3 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

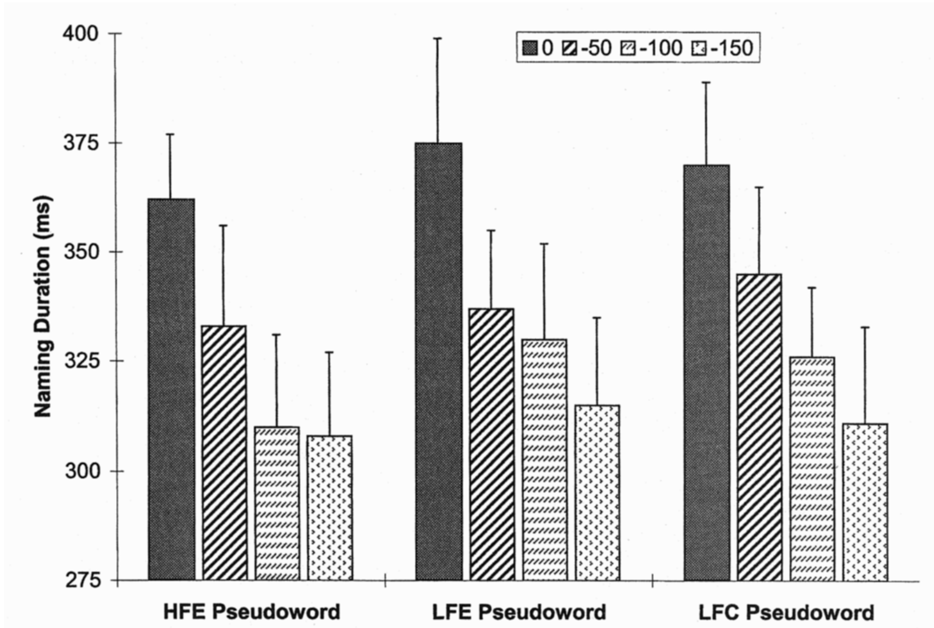

Figure 14. Mean naming durations in the tempo-naming portion of Experiment 3 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

1; B-50 to B-100 was reliable by items but not by subjects, $F_s(1, 31) = 1.9, p > .1, F_i(1, 141) = 3.9, p < .05$; and B-100 to B-150 was fully reliable, $F_s(1, 31) = 5.9, p < .05, F_i(1, 141) = 4.2, p < .05$.

Error counts as a function of error type and tempo are presented in Table 5. To the contrary of the route emphasis hypothesis, the proportion of word errors was greater in the current experiment (53%) than in the previous two (40% in Experiment 1 and 43% in Experiment 2). Furthermore, the rate of increase in word errors as tempo increased was approximately equal to that for nonword errors as was the case in Experiments 1 and 2. The M-H test for trend was significant, $\chi^2(1, N = 391) = 4.8, p < .05$, but inspection of the column percentages shows that this was due to a greater rate of increase in articulatory errors across tempo, compared with word and nonword errors.

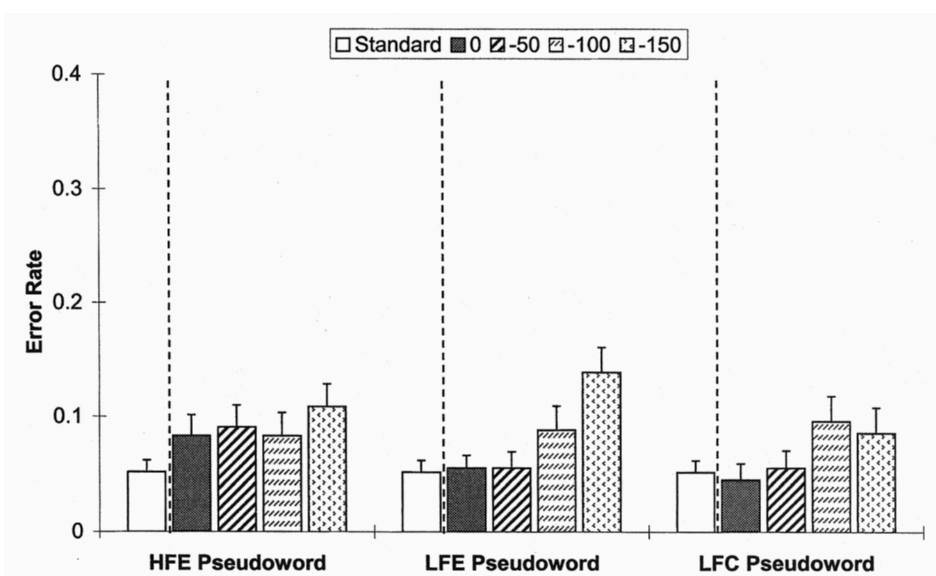

Figure 15. Mean error rates from the standard- and tempo-naming portions of Experiment 3 as a function of stimulus type and tempo. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

Table 5
Frequency Counts of Errors in the Standard- and Tempo-Naming Tasks in Experiment 3, Categorized by Error Type and Tempo (for the Tempo-Naming Task)

		Tempo				
Error type	Standard	0	−50	−100	−150	Total
Word	38	46 (63.0)	41 (51.3)	51 (48.1)	68 (51.5)	206 (52.7)
Nonword	39	23 (31.5)	33 (41.3)	45 (42.5)	43 (32.6)	144 (36.8)
Articulatory	8	4 (5.5)	6 (7.5)	10 (9.4)	21 (15.9)	41 (10.5)
Total	85	73	80	106	132	391

Note. Frequency counts were drawn from 1,600 test responses for standard naming (50 per subject) and 1,152 per tempo block (48 per subject per block). Numbers in parentheses are column percentages.

Discussion

The results of Experiment 3 were consistent with those of Experiments 1 and 2. Timing analyses again suggested that subjects time their responses with the onset of voicing in the tempo-naming task. Furthermore, the number of word, nonword, and articulatory errors as a function of tempo were comparable to those of Experiments 1 and 2: They all steadily increased as tempo was increased. If the route emphasis hypothesis was correct in accounting for the constant number of LARC errors across tempos in Experiments 1 and 2, then one should also expect a decrease in the number of word errors in Experiment 3, which included nothing but pseudowords. The results failed to support this prediction. There are two assumptions to be noted before concluding that route emphasis was not the cause of error patterns in Experiments 1–3.

First, Tabossi and Laghi (1992) have suggested that subjects may be unable to deemphasize the lexical route in English. If this is the case, then Experiment 3 does not bear on the route emphasis explanation. However, even if the lexical route cannot be de-emphasized, one would not expect an increase in the influence of lexical knowledge in Experiment 3 as a function of tempo. We did, in fact, find a significant increase in word errors as a function of tempo, so the study by Tabossi and Laghi would not seem to be an issue. This logic rests on the interpretation of word errors as arising from lexical knowledge, but they may have occurred by chance (e.g., faulty visual or articulatory processing).

We addressed this ambiguity by estimating the rate of chance word errors in Experiment 3. Determining chance error rates is a difficult problem because one must take into consideration the specific stimuli and conditions of the observed errors (e.g., neighborhood density), the types of errors that are possible (e.g., only phonetically legal errors are allowed), as well as any error biases that may be independent of a lexical bias (e.g., in speech production, initial consonants are more likely to produce errors than are final ones; Dell, 1988; Schwartz & Goffman, 1995). We empirically determined a chance error rate for the stimuli in Experiment 3 by using the most conservative method that we could devise given our intentions.

We first estimated the frequency of occurrence of a large number of phonological error types (e.g., position-dependent phoneme substitutions, deletions, insertions, and transpositions) on the basis of the observed errors in Experiments 1–3. We then applied each phonological error type (when applicable), weighted by the frequency estimates, to each test item thereby generating a set of possible error instances. Our estimate of the chance frequency of word errors was the proportion of errors that formed legal English words out of the set of empirically determined possible errors. This method is conservative because the estimated rate of each phonological error type is determined on the basis of all observed errors (i.e., including word errors); these estimates would therefore include any real lexical biases that may correlate with phonological biases (e.g., initial consonant substitutions may cause more word errors than would final consonant substitutions). The chance estimate of the word error rate for Experiment 3 was 28%.[10] The observed rate of word errors was 53%, which was significantly different than chance, $t(27) = 8.1, p < .001$ by subjects, $t(120) = 6.9, p < .001$ by items.

In summary, the assumptions that de-emphasis of the lexical route is possible in English, and that lexical processes contributed to the observed rate of word errors, both seem justified. Therefore, the proportion of word errors observed in Experiment 3 stands as indirect evidence against the route emphasis explanation of the pattern of LARC errors found in Experiments 1 and 2.

General Discussion

We set two goals at the beginning of this study: (a) to formulate a more explicit mechanism of control over response timing and (b) to formulate a hypothesis of how

[10]As a comparison, Garrett (1976) and Dell and Reich (1981) estimated the chance occurrence of word errors in speech production (i.e., within phrases and sentences) as 33% and 40%, respectively. Note that different types of errors can and do occur in speech production versus word naming (e.g., transpositions between words; *barn door* goes to *darn bore*), so one should not necessarily expect these estimates to match.

pressure for speed relates to the time course of processing. The results from Experiments 1–3 can be summarized as follows: Subjects were largely able to entrain response initiation to an external tempo. Stimulus effects on latencies were attenuated in the tempo-naming task compared with the standard-naming task. The tempo manipulation induced naming latencies that were substantially faster and significantly more error prone than were those in the standard-naming task. Naming durations decreased as tempo increased. The rate of LARC errors did not increase with other error types as tempo increased. This result was not attributable to a de-emphasis of the sublexical route. Subjects timed their responses with the onset voicing. This finding is orthogonal to the issues at hand, so we do not consider it further.

The tempo-naming results showed clearly that subjects have some mechanism of response timing available to them. Moreover, the mechanism that controls response initiation seems to be tightly coupled with response execution; the task was to time response initiation with the tempo (and feedback was based on this alone), yet response durations shortened along with latencies. In conjunction with a model of word reading, one might be able to formulate a time criterion mechanism that can account for the latency and error results of the tempo-naming experiments. However, a time criterion alone has no intrinsic implications for response durations, so it cannot account for duration results.

We propose an alternate explanation that is motivated by the evidence for a coupling of response initiation and response execution. Just as the tempo induced compression of the response trajectory, we hypothesize that it also induced compression of the processing trajectory in the word-reading system. Strategic control over response timing does not manipulate the stoppage of the normal course of processing; it changes the course of processing itself such that a response can be initiated at the desired point in time. This property of accelerated processing can be instantiated in a connectionist network by manipulating the input gain to processing units (Cohen & Servan-Schreiber, 1992; Kello, Plaut, & MacWhinney, in press; Nowlan, 1988). Input gain is a multiplicative scaling factor on the net input to processing units (it is equivalent to the inverse of temperature in Boltzmann machines; Ackley, Hinton, & Sejnowski, 1985). The effect of gain on unit outputs depends on their output functions. Consider the common logistic function in a processing unit, a_j, that updates its output in continuous time,

$$a_j^{[t]} = \frac{1}{1 + \exp(-x_j^{[t]}\gamma)}, \qquad (1)$$

in which $x_j^{[t]}$ is the net input to unit a_j at time t, and γ is the input gain. For the logistic, input gain serves to sharpen (for large values of γ) or flatten (for small values of γ) the effect that a change in the net input has on the output. For high values of gain, small changes in the net input to a unit can be sufficient to move the output between 0 and 1. For low values of gain, large changes in the net input are necessary to have a comparable effect on a unit's output.

Raising the input gain on units in a network can accelerate processing because net inputs can change unit outputs in a smaller number of time steps. With a nonlinear activation function and interactions among processing units, the modulation of gain can reverse the asymptotic states of units. Depending on how unit states are interpreted, this can correspond to a qualitative change in network behavior. Kello et al. (in press) have demonstrated both of these basic effects in an abstract connectionist model of information processing.

In addition to the coupling of response latencies and durations, how might the manipulation of input gain in a connectionist model of word reading account for the other findings from the tempo-naming experiments? The two most relevant results to consider are (a) the attenuation of stimulus effects on latencies and (b) the pattern of errors. The first falls naturally out of the fact that accelerated processes are compressed in time. Other factors being equal, as overall processing times are shortened, any differences in processing times across conditions should also be shortened.

With regard to the error results, there are two findings to account for: (a) the overall increase in error rate with increased tempos and (b) the corresponding increase in the rate of all error types except LARC errors. First, Kello et al. (in press) have shown that high levels of gain can cause a general loss in accuracy. This can occur if increased gain amplifies any noise in processing or if the relative timing of unit output trajectories is disturbed.

With regard to the second finding, let us consider the rate of occurrence of each error type as a function of tempo. In all three experiments, the rate of word, nonword, and articulatory errors all increased with faster tempos, whereas the rate of LARC errors (in Experiments 1 and 2) remained constant. We assume that input gain is being manipulated across all processing units in a model of word reading. This is a logical extension of our hypothesis that internal processes are coupled with motor processes. Given this, it is straightforward to explain the increase in nonword and articulatory errors. A primary source of articulatory errors is likely to be within the processes that map phonological representations to motor commands. As gain is increased within these processes, the rate of articulatory errors should increase. Analogously, a primary source of nonword errors is likely to be within phonological processes (e.g., phonotactic constraints), so as gain is increased within phonology, the rate of nonword errors should increase.

By this logic, a primary source of word and LARC errors would be a dysfunction in the mapping from orthography to phonology (although phonological and orthographic–visual processes may also contribute). The above-chance increase in word errors, coupled with the constant rate of LARC errors, suggests that increased tempos caused a proportional amplification of the influence of lexical knowledge versus sublexical knowledge. How might an increase in the level of gain in a connectionist model of word reading cause such an effect? In the triangle framework, lexical knowledge has its influence primarily through semantics, whereas sublexical knowledge is stored primarily in the weights between orthography and phonology. Therefore, to account for the constant rate of LARC errors, an increase in input gain would have to emphasize the contribution of semantics over sublexical knowledge.

We hypothesize that gain would have this effect because of the arbitrary nature of the relationship between semantics and phonology relative to the systematic relationship between orthography and phonology. Overall, networks have a tendency to map similar inputs to similar outputs. This property facilitates the mapping from orthography to phonology because of the systematicity in their relationship, and therefore the weights between these two levels of representation do not have to grow very large during learning. By contrast, the nonsystematic relationship between semantics and phonology requires larger weights to overcome the bias for similar inputs to produce similar outputs. Input gain is multiplicative with respect to weight magnitude, but the contributions from different incoming weights are additive with respect to a receiving unit's net input. Therefore, an increase in gain will amplify the influence of larger weights over smaller weights. Given that weights from semantics to phonology are larger than those from orthography to phonology, one might expect an increase in input gain to amplify lexical knowledge over sublexical knowledge.

It remains to be seen whether in a full-scale model of word reading the manipulation of input gain can be demonstrated to have the aforementioned properties. Preliminary simulations have been encouraging (Kello & Plaut, 1998), and a full treatment of these issues is the topic of ongoing work (Kello & Plaut, 1999).

Stimulus Blocking and Input Gain

Part of the motivation for the tempo-naming experiments was the proposal of a time criterion to account for some effects of stimulus blocking found by Lupker, Brown, et al. (1997) and Jared (1997). We have argued that it is difficult to account for the tempo-naming results with a time criterion, and we have proposed the gain hypothesis as an alternative. To what extent can the gain hypothesis account for stimulus-blocking effects?

As a mechanism to account for stimulus-blocking effects, input gain is actually in the same spirit as a time criterion. Both accounts are distinguished from the route emphasis hypothesis in that they each argue for strategic control over response initiation rather than the processing of individual routes. Furthermore, both hypotheses propose that a response criterion is shifted as a function of stimulus difficulty; as stimuli in a block become more difficult, on average, the response criterion will shift to a more conservative setting. As explained previously, the key difference between these two hypotheses is the exact mechanism underlying the response criterion. The time criterion halts the normal trajectory of processing at a particular point in time, whereas input gain accelerates or decelerates the trajectory of processing.

This differentiates the two hypotheses in how predictions are made on the basis of latency data in stimulus-blocking experiments. Lupker, Brown, et al. (1997) and Jared (1997) used mean latencies in the pure blocks conditions as a direct measure of the relative positioning of the time criterion, and they used this measure to predict mean latencies in the mixed blocks. For example, Lupker, Brown, et al. (1997) found that the mean latency of a pure block of high-frequency consistent (HFC) words was faster than that of a pure block of pseudowords. They interpreted this difference in mean latencies as indicating that the time criterion was set more conservatively in the pseudoword block than it was in the HFC block. The time criterion hypothesis stipulates that a middling time criterion should be set in a mixed block of both pseudowords and HFC words. Therefore, the hypothesis predicts an increase in HFC latencies and a decrease in pseudoword latencies in the mixed block; this is what they found.

The gain hypothesis can make the same prediction, but one must interpret the pure block latencies in the context of a theory of the mapping from orthography to phonology. This is because latencies are a function of gain in conjunction with stimulus type, and the effect of stimulus type on latencies is determined (at least in part) by one's theory of how orthography is mapped to phonology. Strictly speaking, this is also true of the time criterion hypothesis. As discussed in the introduction, latencies cannot be a function of the time criterion alone because this would predict no stimulus effects. However, the contribution of the time criterion to latencies is somewhat independent of the mapping from orthography to phonology, whereas gain exerts its influence on latencies through this mapping. The current study does not make claims about how orthography is mapped to phonology (i.e., we are not presenting a theory of word reading here), so we cannot quantitatively determine how well the gain hypothesis can account for the relevant stimulus-blocking effects. However, qualitatively speaking, gain is very similar to a time criterion as a mechanism of strategic control, so it will have a similar set of problems and benefits.

Other Uses of Input Gain

Input gain has been used as a mechanism of strategic control in other studies as well. In one line of research, gain has been used as a mechanism of control over the influence of contextual information on stimulus processing (Cohen & Servan-Schreiber, 1992). In that study, a connectionist model of Stroop phenomena was presented in which the role of two processing units was to provide task information (context, i.e., one unit represented the color-naming task and the other represented the word-naming task). The input gain of the task units (mathematically equivalent to the gain parameter used in the current study) was manipulated to simulate the hypothesized role of the neurotransmitter dopamine in the prefrontal cortex (PFC). A large body of neurophysiological evidence has indicated that dopamine may modulate the gain of postsynaptic input summation in the PFC (as well as in other areas; see Cohen & Servan-Schreiber, 1992), and Cohen and Servan-Schreiber view the function of the PFC as the maintenance of task and situation context. Furthermore, other studies have shown that the regulation of dopamine is impaired in schizophrenics (Cohen & Servan-Schreiber, 1992). Therefore, the modulation of gain in their model served to simulate normal versus schizophrenic performance in Stroop and related tasks.

In a study more similar to the current one, Kello et al. (in press) conducted two Stroop color-naming experiments to examine the temporal relationship between cognitive processing and overt articulation. They found that as subjects were pressured to respond faster (by the introduction of a deadline), the effect of Stroop interference bled over from naming latencies to naming durations. Just as in the current study, input gain was invoked as a mechanism to strategically control the speed of response initiation. The authors presented an abstract connectionist model of information processing that captured the basic effect of the deadline in their experiments by manipulating the input gain to processing units in the network. Relevant to the tempo-naming experiments, the authors found that an increase in gain caused both response latencies and durations to shorten in the network. The efficacy of input gain as a mechanism of control over response speed in the study by Kello et al. is evidence that input gain may account for the tempo results and may develop into a more general theory of cognitive control.

The current study and the Kello et al. (in press) study invoked gain as a mechanism of control over processing speed, whereas the Cohen and Servan-Schreiber (1992) study invoked gain to modulate the influence of certain information on stimulus processing. These are different interpretations of the specific role that gain plays, but they both treat gain as a mechanism of strategic control.

Conclusions

We introduced the tempo-naming task to investigate mechanisms of response timing and to provide a new empirical window into the time course of phonological processing in word reading. The results of three tempo-naming experiments were interpreted as evidence against the use of a time criterion as the mechanism of response timing in tempo naming. With regard to the time course of phonological processing, fast tempos caused an increase in all error types except LARC errors. This result was interpreted as evidence that the influence of sublexical knowledge on the mapping from orthography to phonology is reduced under pressure for speeded responding. We proposed that input gain, as a mechanism of control over processing speed in the word-reading system, can potentially account for the tempo-naming results. Whether input gain can provide a general theory of strategic control over processing is a topic for future studies.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann Machines. *Cognitive Science, 9,* 147–169.

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

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language, 28,* 14–36.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24,* 89–106.

Berndt, R. S., Reggia, J. A., & Mitchum, C. C. (1987). Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments, & Computers, 19,* 1–9.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3,* 395–421.

Cody, R. P., & Smith, J. K. (1997). *Applied statistics and the SAS programming language.* Upper Saddle River, NJ: Prentice-Hall.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99,* 45–77.

Colombo, L., & Tabossi, P. (1992). Strategies and stress assignment: Evidence from a shallow orthography. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 319–399). Amsterdam: North-Holland.

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.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629.

Forster, K. I., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30,* 1–25.

Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361–379.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 133–177). Amsterdam: North-Holland.

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

Harm, M. W. (1998). *Division of labor in a computational model of visual word recognition.* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Herdman, C. M. (1992). Attentional resource demands of visual word recognition in naming and lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 460–470.

Herdman, C. M., LeFevre, J. A., & Greenham, S. L. (1996). Base-word frequency and pseudohomophone naming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A,* 1044–1061.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124,* 62–82.

Hoequist, C. E. (1983). The perceptual center and rhythm categories. *Language and Speech, 26,* 367–376.

Jared, D. (1997). Evidence that strategy effects in word naming reflect changes in output timing rather than changes in processing route. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1424–1438.

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

Kawamoto, A. H. (1988). Distributed representations of ambiguous words and their resolution in a connectionist network. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 195–228). San Mateo, CA: Morgan Kaufmann.

Kawamoto, A. H., Kello, C. T., Higareda, I., & Vu, J. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 362–381.

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 862–885.

Kawamoto, A. H., & Kitzis, S. N. (1991). Time course of regular and irregular pronunciations. *Connection Science, 3,* 207–217.

Kello, C. T., & Kawamoto, A. H. (1998). Runword: An IBM-PC software package for the collection and acoustic analysis of speeded naming responses. *Behavior Research Methods, Instruments, and Computers, 30,* 371–383.

Kello, C. T., & Plaut, D. C. (1998, November). *Modulation of lexical effects in word reading without strategic control of pathways: Evidence from tempo naming.* Paper presented at the 39th Annual Meeting of the Psychonomic Society, Dallas, TX.

Kello, C. T., & Plaut, D. C. (1999, November). *Strategic control by gain manipulation in a computational model of word reading in the tempo naming task.* Paper presented at the 40th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (in press). *The task-dependence of staged versus cascaded processing: An empirical and computational study of stroop interference in speech production. Journal of Experimental Psychology: General.*

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

Kurganskii, A. V. (1994). Quantitative analysis of periodical time-patterned finger tapping. *Human Physiology, 20,* 117–120.

Lemoine, H. E., Levy, B. A., & Hutchinson, A. (1993). Increasing the naming speed of poor readers: Representations formed across repetitions. *Journal of Experimental Child Psychology, 55,* 297–328.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 570–590.

Lupker, S. J., Taylor, T. E., & Pexman, P. M. (1997, November). *Strategic control of a time criterion in naming: New evidence and effects.* Paper presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.

Manis, F. R. (1985). Acquisition of word identification skills in normal and disabled readers. *Journal of Educational Psychology, 77,* 78–90.

Marcus, S. M. (1981). Acoustic determinants of perceptual center (p-center) location. *Perception and Psychophysics, 52,* 691–704.

Mates, J., Radil, T., & Poppel, E. (1992). Cooperative tapping: Time control under different feedback conditions. *Perception and Psychophysics, 52,* 691–704.

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

McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language, 29,* 43–65.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed–accuracy decomposition. *Psychological Review, 95,* 183–237.

Monsell, S., Patterson, K., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452–467.

Nowlan, S. J. (1988). Gain variation in recurrent error propagation networks. *Complex Systems, 2,* 305–320.

Ollman, R. T., & Billington, M. J. (1972). The deadline model for simple reaction times. *Cognitive Psychology, 3,* 311–336.

Orden, G. C. V. (1991). Phonological mediation is fundamental in reading. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 77–103). Hillsdale, NJ: Erlbaum.

Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research, 53,* 13–24.

Pachella, R. G., & Pew, R. W. (1968). Speed–accuracy tradeoff in reaction time: Effect of discrete criterion times. *Journal of Experimental Psychology, 76,* 19–24.

Patterson, K., & Behrmann, M. (1997). Frequency and consistency effects in a pure surface dyslexic patient. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1217–1231.

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

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 482–503.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Ruthruff, E. (1996). A test of the deadline model for speed–accuracy tradeoffs. *Perception and Psychophysics, 58,* 56–64.

Savage, G. R., Bradley, D. C., & Forster, K. I. (1990). Word frequency and the pronunciation task: The contribution of articulatory fluency. *Language and Cognitive Processes, 5,* 203–236.

Schwartz, R. G., & Goffman, L. (1995). Metrical patterns of words and production accuracy. *Journal of Speech and Hearing Research, 96,* 523–568.

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

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L.,

& McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1177–1196.

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.

Sherak, R. (1982). A real-time software voice key and an application. *Behavior Research Methods and Instrumentation, 14,* 124–127.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition, 6,* 410–415.

Stanovich, K. E., & Pachella, R. G. (1976). The effect of stimulus probability on the speed and accuracy of naming alphanumeric stimuli. *Bulletin of the Psychonomic Society, 8,* 281–284.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89,* 114–127.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 507–534). Hillsdale, NJ: Erlbaum.

Strain, E., Patterson, K., Graham, N., & Hodges, J. R. (1998). Word reading in Alzheimer's disease: Cross-sectional and longitudinal analyses of response time and accuracy data. *Neuropsychologia, 36,* 155–171.

Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: II. Dynamic aspects of strategy adjustment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 342–365.

Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition, 20,* 303–313.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language, 26,* 608–631.

Treisman, A., & Williams, T. C. (1991). A theory of criteria setting with an application to sequential dependencies. *Psychological Review, 91,* 68–111.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. T. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 528–553.

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.

Venezky, R. L. (1970). *The structure of English orthography.* The Hague, the Netherlands: Mouton.

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

Wickelgren, W. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

Yamada, N. (1995). Nature of variability in rhythmical movement. *Human Movement Science, 14,* 371–384.

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

Appendix A

Stimuli From Experiments 1 and 2

Standard naming				Tempo naming			
HFE	LFE	LFC	Pseudo	HFE	LFE	LFC	Pseudo
break	breast	brunt	brish	blood	blown	bloke	blisp
both	bush	bum	bax	both	bush	bum	bift
friend	frost	fret	frope	break	breast	brunt	brote
full	flange	flask	flade	broad	plaid	plod	plake
gone	guise	grope	gret	come	caste	cask	cet
have	hearth	hark	hoam	cost	comb	coal	cipe
low	limb	lisp	lask	dead	dough	dole	dest
put	pear	pest	pag	death	deaf	desk	detch
source	suede	swish	swench	done	drought	dank	dit
want	wand	wit	weke	door	doll	dolt	doan
what	wad	wax	wub	foot	flood	flake	flain
where	wool	wade	wibe	four	flown	float	flill
word	womb	wane	wuff	friend	frost	fret	frask
are	ere	eke	ean	full	flange	flask	flod
dog	cache	crag	crelp	give	glove	glum	glane
his	swath	swoon	swunt	gone	guise	grope	grole
hour	rouge	rune	roon	good	gong	gob	gade
none	mule	mend	mest	great	grown	groan	grig
once	wharf	whelp	whark	gross	ghoul	goon	gax
there	chic	shrub	shrane	have	hearth	hark	hait
warm	wolf	wilt	wune	head	hind	hitch	hame
watch	rouse	roam	rilt	heard	hook	hump	hink
whom	vise	vibe	vit	key	cough	cuff	coom
whose	farce	fluff	flisp	learn	leapt	letch	lesk
worth	swamp	stench	stend	lost	loath	lobe	loke
your	wan	wean	wum	love	lose	loom	luff
				low	limb	lisp	lum
				month	mourn	munch	meep
				most	mould	mole	mell
				move	mow	moan	mick
				poor	pint	pine	pank
				put	pear	pest	pite
				said	swap	swig	swask
				says	spook	spurt	spole
				shall	shoe	shame	shump
				show	shove	shell	shunch
				some	sew	sole	surt
				son	steak	stain	stine
				source	suede	swish	swobe
				stood	suave	swill	swope
				through	threat	thrift	thrish
				touch	tomb	tote	tunt
				truth	trough	trite	tritch
				two	tread	trait	troat
				want	wand	wit	wark
				war	wart	wink	woal
				were	worm	wick	woan
				what	wad	wax	wob
				where	wool	wade	wole
				word	womb	wane	wolt
				work	wasp	wipe	woon
				would	warp	weep	wum

Note. Pseudowords only appeared in Experiment 2. *Chic* was removed from the analyses because of an inordinate number of errors. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent; Pseudo = pseudowords.

Appendix B

Stimuli From Experiment 3

	Tempo				Tempo		
Standard	HFE	LFE	LFC	Standard	HFE	LFE	LFC
swant	bloor	blad	blift	losh	mave	mough	murt
wug	bood	bost	bope	loup	mork	pleaf	pake
blan	breat	broul	brole	mape	pead	pove	pank
blig	brone	cearth	cark	meeb	pone	pown	shain
boog	cays	cong	cobe	morp	pood	sarp	shig
brear	diend	cood	dest	murp	sall	shart	spoan
carg	dord	deak	disp	pleam	seard	shint	stell
cowe	dource	deast	dit	poot	shouch	sook	sunt
dait	dut	duave	flane	sak	sood	sourn	swax
dier	fost	flaste	fletch	shap	sove	spown	swole
diz	foth	flomb	fline	shink	stost	swand	tait
dorg	fough	flove	fritch	soop	sull	thromb	throke
dup	frome	frap	gite	soor	throot	torm	trob
flet	gaid	geat	glick	spow	tove	trind	trote
flote	gead	ghough	goom	sunth	trour	trould	wame
frak	goor	glead	grum	swode	twey	woll	wask
gabe	grat	goath	hade	throob	wearn	wook	woat
geave	grive	grear	het	tring	woad	wose	wole
gleap	har	haid	hink	trouch	wome	wough	woon
grap	heak	heapt	hod	wape	wonth	wought	wum
gream	huth	huede	lask	woap	woss	wown	wump
grud	ko	lange	lish	wom	wost	wush	wunch
heab	leath	lew	lolt	woop			
hib	lon	lomb	luff	wosk			
huf	lould	masp	meep	woup			
leab	mant	mool	mipe	wunt			

Note. HFE = high-frequency exception; LFE = low-frequency exception; LFC = low-frequency consistent.

Received August 21, 1998
Revision received October 5, 1999
Accepted October 12, 1999 ■