Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 5, 1289-1302

0278-7393/95/$3.00

Mental Chronometry: Beyond Onset Latencies in the Lexical Decision Task

David A. Balota and Richard A. Abrams
Washington University

Four experiments addressed the influence of variables on lexical decision performance *after* response initiation has occurred. In Experiment 1, participants made an arm movement in one direction for word trials and pressed a button with the other hand for nonword trials. The results indicated that word frequency not only modulated the speed to initiate the arm movement, but also modulated the acceleration and force of the movement after initiation. The results of Experiment 2 indicated that word frequency and stimulus degradation produced large and additive effects on response latency, accuracy, and the force of the response after initiation. In Experiments 3 and 4, participants made the *same* arbitrary speech response in a modified lexical decision task for both high- and low-frequency words. The results indicated that both the onset and duration of the speech response were modulated by word frequency. The results are viewed as most consistent with an enabled response model, wherein early operations can enable appropriate action systems before central decisions are made.

Mental chronometry is the study of the time course of information processing in the human nervous system (Posner, 1978). Although there have been many analytic techniques developed within the mental chronometry enterprise (see Meyer, Osman, Irwin, & Yantis, 1988, for a review), the vast majority of this research effort has relied on reaction time (RT) techniques. Typically, during an RT trial, a stimulus is presented and the research participant is required to make a judgment as quickly and as accurately as possible regarding some dimension or dimensions of the stimulus. Participants usually indicate their decision by pressing one of several possible response buttons or, in some cases, producing a simple vocal response. If an experimental factor influences the time taken to initiate the response, then it is generally assumed that the factor influences processes prior to or including the central decision process, but not processes after the decision has been made. Presumably, once a central decision has been made and the response has been initiated, the influence of the experimental factor is over. As a number of researchers have noted (e.g., Luce, 1986; Ulrich & Wing, 1991), the major reason for this assumption is that it greatly simplifies the development of models of RT performance. Thus, researchers within the mental chronometry tradition rarely measure the characteristics of a response beyond the triggering of a microswitch.

RT experiments have provided a wealth of knowledge about details of human information processing. However, experiments in which RT is the primary (or sole) dependent measure are limited: They provide a measure of the temporal requirements of various types of decisions—but only at a single point in the information-processing stream. Other events before and/or after response onset may be equally, or perhaps even more, revealing of aspects of the mental operations that are taking place. For example, consider the possibility that early in processing, Factor A produces a plus 10-ms effect, whereas later in processing, Factor A produces a minus 10-ms effect, possibly by means of a reciprocal inhibitory process. The null effect of Factor A in an RT experiment would be quite misleading. In fact, one of the major reasons for the recent enthusiasm for the use of electrophysiological measures (e.g., event related potentials and electromyogram) is that such measures allow one to track the influence of factors in a more continuous fashion. In fact, the results from electrophysiological studies clearly indicate that conclusions based exclusively on effects at the moment of response onset are limited. Consider, for example, the research of Kutas, McCarthy, and Donchin (1977). They found that the P300 component of the evoked response can sometimes be shorter, longer, or equal to RT differences across different conditions. Coles, Gratton, Bashore, Eriksen, and Donchin (1985) have persuasively argued that the evoked response technology avoids many of the inherent problems with the single-point, traditional estimates of information flow that RT paradigms provide.

In addition to the possibility of learning about processes by continuously sampling activity during relatively early stimulus evaluation processes (as is possible with electrophysiological techniques), it is also possible that one may be able to learn more about mental activity by continuous sampling of behavior *after* response initiation has occurred. There are a number of reasons to believe that events after the onset of the response might be informative. For example, work by Osman, Kornblum, and Meyer (1986) suggests that some experimental

---

Portions of this work were presented at the 32nd Annual Meeting of the Psychonomic Society, San Francisco, California. The work was supported by National Institute of Aging Grants PO1 AGO3991 and RO1 AG10193 and by National Institutes of Health Grant R29-MH45145. We thank Derek Besner for suggesting the degradation manipulation used in Experiment 2 and Mark Faust, James Neely, and Richard Schweickert for helpful comments on a draft of this article. This article represents a collaborative effort.

Correspondence concerning this article should be addressed to either David A. Balota or Richard A. Abrams, Department of Psychology, Box 1125, Washington University, St. Louis, Missouri 63130. Electronic mail may be sent via Internet to dbalota@artsci.wustl.edu.

factors may influence performance after response selection has occurred (i.e., very late in the information processing stream—after the point of no return). In addition there is already some evidence available in the literature that suggests that some variables can influence the characteristics of a response after response initiation has taken place. For example, Balota, Boland, and Shields (1989) have provided evidence that not only are the onset latencies to produce a word (e.g., *dog*) shorter when it is primed by a related word (e.g., *cat*) compared to an unrelated word (e.g., *pen*), but under some conditions, the production durations of the word (i.e., from the first phonetic segment to the last phonetic segment) are also shorter (see also Balota & Shields, 1988; Fowler, 1988; Fowler & Housum, 1987; Lieberman, 1963; Shields & Balota, 1991). In addition, Coles et al. (1985) have provided evidence that the presence of an incompatible flanking stimulus (e.g., SS*H*SS vs. HH*H*HH, where the central letter is the target to which participants responded) not only influenced the time taken to initiate a correct response but also influenced more peripheral aspects of the response (e.g., the interval between electromyographic activity in the correct response channel and the onset of the overt correct response). Finally, Giray and Ulrich (1993) have recently demonstrated that bimodal (auditory and visual) redundant stimuli produced both decreases in response latency and increases in response force, compared with unimodal stimuli, that is, produced a redundant-signals effect in both response latency and force measures. These authors argued that their results challenged the common assumption that early processes do not influence later motoric aspects of a response after initiation has occurred.

We (Abrams & Balota, 1991) reported evidence of the continuous nature of information flow after response initiation by investigating simple limb movements in two classic RT paradigms: lexical decision and memory scanning. At the beginning of each trial within these experiments, participants positioned a handle in a central location. They were required to make a rapid limb movement in one of two opposing directions to reflect their binary decision in each task. For example, in the lexical decision task, participants might move the handle in the rightward direction to respond to a word and in the leftward direction to respond to a nonword. In addition to the lexicality of the stimulus, we also manipulated a variable, word frequency, that presumably influences factors leading up to the central (word–nonword) decision process. As expected, the results indicated that high-frequency words produced faster onset latencies (initiation of the limb movements) than did low-frequency words. More interesting, however, was the finding that word frequency also affected kinematic aspects of the responses, after they had been initiated. Specifically, responses to high-frequency words were more forceful than were responses to low-frequency words. If word frequency primarily influences the time taken to recognize a target word (i.e., processes leading up to the central decision about the lexicality of the stimulus), and the response system blindly executes the response, then one would not expect such influences of word frequency on response dynamics after response initiation. Similar results were also found in the memory scanning paradigm, wherein both onset latencies and response dynamics after response initiation were influenced by memory-set size and by the presence of the probe in the memory set (Abrams & Balota, 1991, Experiment 2).

The present set of experiments was designed to further explore the influence of word frequency in the lexical decision task beyond response initiation. We chose to examine the nature of the word-frequency effect in the lexical decision task because (a) the lexical decision task is one of the most widely used tasks in the RT literature, and (b) the word-frequency effect has played a primary role in virtually all models of word recognition. Moreover, there has been considerable recent discussion regarding how this variable influences performance in this task (e.g., Balota & Chumbley, 1984, 1990; Monsell, 1990; Monsell, Doyle, & Haggard, 1989). As discussed below, the present results will shed some light on how word frequency modulates performance in the lexical decision task.

In Experiment 1, we first attempted to replicate our original word-frequency effect on response force (Abrams & Balota, 1991), while at the same time eliminating an alternative, and quite plausible, response competition interpretation of the earlier findings. Experiment 2 included a manipulation of stimulus quality in addition to the frequency manipulation to determine if factors other than frequency influence response characteristics in lexical decision performance. In particular, the goal of Experiment 2 was to determine whether a factor that has a relatively early, and independent influence on lexical processing, stimulus degradation, will also influence response force. Experiments 3 and 4 involved a modified lexical decision task and were designed to determine whether such effects would extend to a different response modality, speech, by having participants produce the same vocal response (*normal* in Experiment 3 and, *berloe* in Experiment 4) to both high-frequency and low-frequency words. Experiments 3 and 4 are particularly relevant to current models of speech production that emphasize activation and selection of phonological elements in speech production and place little emphasis on the execution of these accessed elements. The results of these experiments indicate that variables that influence activation and selection can also influence execution processes in fluent speech productions.

Experiment 1

As noted earlier, Abrams and Balota (1991) reported evidence that limb movements in response to high-frequency words were not only initiated faster than in response to low-frequency words, but they were also more forceful than responses to low-frequency words. In Experiment 1, we examined a plausible alternative explanation of this finding. Specifically, in the Abrams and Balota study, participants judged the lexical status of a letter string by moving a handle in one direction if the letter string was a word and by moving the same handle in the opposite direction if the letter string was a nonword. Because participants moved the same handle in opposing directions for word and nonword responses, it is possible that the observed effect of word frequency on force may have been due to competition between the two possible responses. Consider the possibility that, as some have argued (e.g., Balota & Chumbley, 1984; Besner & Swan, 1982), low-frequency words are more similar to nonwords on a dimension (familiarity) that is relevant to the demands of the lexical decision task. By having participants move the same handle in one direction for a word response and a different direction for a nonword response, there may have been peripheral response competition (pulling the handle in opposing directions) that was stronger for low-frequency words than for high-frequency words. This would decrease the force of the response to low-frequency words compared with the force of the response to high-frequency words, which is the result that was observed. Of course, in the typical RT paradigm, participants make responses that are not mutually exclusive (e.g., keypresses involving the two index fingers). Hence, it is possible that the effect of word-frequency on response force observed in our earlier study occurs in the lexical decision task only when participants make physically opposing responses with the same limb. In Experiment 1, we tested this possibility by having participants make a limb movement with one hand on word trials, whereas for nonword trials, participants simply pressed a key with the other hand. Because two separate limbs were involved for word and nonword responses, any observed word-frequency effect could not be due to this type of response competition.

Method

Participants. Twenty right-handed undergraduate volunteers from the Washington University community participated in this experiment. They were all native English speakers and naive to the purposes of the experiment and each was paid $5 for participating. No participant was involved in more than one of the present experiments.

Equipment. The presentation of stimuli and acquisition of data throughout the experiment was controlled by an IBM-compatible microcomputer. Participants were seated at a table in front of a monitor driven by the computer. Viewing distance was about 40 cm, at which approximately three letters on the CRT subtended 1° of visual angle. Each participant rested their right elbow on the table and grasped a modified joystick handle in their right hand. The handle was mounted directly in front of the monitor on a metal rail fitted with ball-bearings that allowed smooth movement in one dimension along a line parallel to the face of the monitor. The position of the handle was measured by a transducer that produced a position-dependent voltage that was sampled and digitized at 1 kHz with a resolution of .01 cm.

Materials. Thirty-six high- and 36 low-frequency target words were selected from Kučera and Francis's (1967) norms. The low-frequency words had counts less than 7 per million while the high-frequency words had counts greater than 36 per million. The high- and low-frequency words along with 72 pronounceable nonwords were matched on number of syllables, phonemes, and letters. There were also 23 words and 23 pronounceable nonwords that served as practice and buffer items.

Procedure. At the beginning of each trial a small dot appeared at the center of the monitor, and two small squares were displayed 8° (of visual angle) to the left and right of the dot. The right square served as the "target" for the movement—marking the location past which the participants would be required to move the handle on word trials; the left square was included to maintain compatibility with our previous experiments (see Abrams & Balota, 1991). The participant aligned the handle to the center dot with the aid of auditory signals. After the handle was aligned, the dot changed to a plus sign which served as a warning signal. One second later, a letter string was presented. If the string was a word, participants were required to rapidly move the handle to the right. They were instructed that the response would be completed only when the handle passed the right square (which required moving the handle 5.3 cm), at which time a brief tone was presented. To ensure that participants were not simply stopping the handle immediately past the target area, they were encouraged to move the handle well beyond the target point. On nonword trials, participants were instructed to press the space-bar on the computer keyboard with their left index finger.

Participants first received a block of 38 practice trials. The experimenter remained with the participant during the first few trials to ensure that he or she fully understood the instructions. After the practice block, participants served in two blocks of 76 trials each, with the first 4 trials in each block being buffer trials (data from the buffer trials were not analyzed). Participants received a short break after the practice block and after the first test block. All participants were tested individually and their performance was monitored by means of a control station in an adjacent room.

Movement analysis. The handle position was digitally differentiated and filtered to obtain smoothed records of velocity and acceleration as a function of time (see Meyer, Abrams, Kornblum, Wright, & Smith, 1988, for full details). Movement onset was defined as the first moment in time when the velocity of the handle exceeded 3.3 cm/s and remained above that value for at least 20 ms. The end of movement was defined to be the first moment at which the handle passed the target location. These are the same criteria used by Abrams and Balota (1991) to define a movement.

Results

Mean response latencies, accuracies, and kinematic features of the responses are displayed in Table 1 as a function of word frequency. Errors included anticipations (RTs less than 100 ms), late responses (key press RTs greater than 1,200 ms or movement onsets greater than 1,000 ms), and incorrect keypresses on word trials. For each of the following experiments, we report F-statistics from two separate analyses: a subjects-based analysis of variance (ANOVA), referred to as F_1, and an items-based ANOVA, referred to as F_2.

As expected, responses were initiated sooner, $F_1(1, 19) = 498.75$, $MSE = 400.42$, $p < .0001$; $F_2(1, 70) = 235.42$, $MSE = 1390$, $p < .001$, for high-frequency words compared with

Table 1

Mean Response Latencies, Accuracies, and Characteristics of the Response After Response Initiation for Experiment 1 as a Function of Word Frequency

Dependent variable	Word frequency	
	High	Low
Response latency (ms)^{a}	495.1	636.4
% correct^{a}	98	78
Characteristics		
Movement duration (ms)^{a}	127.4	133.9
End velocity^{a} (cm/s)	115.7	108.0
Peak acceleration^{a} (cm/s/s)	192.1	175.3
20-ms after onset		
Position (cm)	.194	.218
Velocity (cm/s)	19.98	19.09
Acceleration^{a} (cm/s/s)	112.23	102.00
50-ms after onset		
Position (cm)	1.39	1.33
Velocity^{a} (cm/s)	60.50	55.57
Acceleration^{a} (cm/s/s)	148.04	132.08

^{a}A reliable frequency effect at least at the $p < .05$ level.

low-frequency words. The effect of word frequency on response latencies corresponded to the effect of word frequency on button press latencies in the traditional lexical decision task and presumably reflected the time taken to make a decision regarding the lexicality of the stimulus. More important, as shown in Table 1, compared with low-frequency words, high-frequency words produced (a) shorter total movement durations, $F_1(1, 19) = 18.62$, $MSE = 22.64$, $p < .001$; $F_2(1, 70) = 18.18$, $MSE = 102.91$, $p < .001$; (b) faster velocities at the end of the movements, $F_1(1, 19) = 16.46$, $MSE = 36.50$, $p < .001$; $F_2(1, 70) = 22.37$, $MSE = 43.12$, $p < .001$; and (c) greater peak accelerations during the movements, $F_1(1, 19) = 12.56$, $MSE = 223.75$, $p = .002$; $F_2(1, 70) = 24.78$, $MSE = 259.16$, $p < .001$. Moreover, if one considers the movements at two different points in time (viz., 20 ms and 50 ms after response initiation), one can see that at 20 ms, the acceleration of the movement was greater for high-frequency words than for low-frequency words, $F_1(1, 19) = 19.84$, $MSE = 58.84$, $p < .001$; $F_1(1, 70) = 14.33$, $MSE = 138.67$, $p < .001$, and that by 50 ms after movement onset, there was still a difference in acceleration, $F_1(1, 19) = 14.43$, $MSE = 176.44$, $p = .001$; $F_2(1, 70) = 38.79$, $MSE = 167.15$, $p < .001$, along with a reliable difference in velocity, $F_1(1, 19) = 15.69$, $MSE = 15.44$, $p = .001$; $F_2(1, 70) = 20.78$, $MSE = 259.16$, $p < .001$. Therefore, when one considers the dynamics of the response after response initiation, high-frequency words produced more forceful responses than did low-frequency words.

Discussion

The results of Experiment 1 revealed a reliable influence of word frequency on both the initiation of the response and the force of the response after initiation. This pattern was obtained even though participants did not make opposing word and nonword responses with the same limb. Thus, we are able to rule out the response competition explanation of the earlier Abrams and Balota (1991) study. Of course, it is possible that the effect was reduced somewhat in the present study compared to the previous study, suggesting at least some role for a more peripheral response competition effect. However, this does not appear to be the case because the effect of word frequency on response dynamics was actually slightly stronger in the present study. For example, the effect of word frequency on movement duration in the Abrams and Balota study was 3.8 ms, whereas, in the present study it was 6.5 ms. This clearly would suggest that the type of responses selected for the binary lexical decisions in our earlier study was not modulating the obtained frequency effects.

Experiment 2

In Experiment 2, we investigated the conjoint influence of stimulus degradation and word frequency to determine whether the response force measures would also be sensitive to a factor that has a relatively early influence in the information processing flow. The motivation for this manipulation was based on the considerable evidence that visual degradation (compared with word frequency) influences an earlier and separate stage of information processing. For example, Becker and Killion (1977), Borowsky and Besner (1993), Schuberth, Spoehr, and Lane (1981), Stanners, Jastrzembski, and Westbrook (1975), and Wilding (1988, Experiment 2) have all reported evidence that visual degradation and word frequency produce additive effects in the lexical decision task (see, however, Norris, 1984; Wilding, 1988, Experiment 1). On the basis of additive-factors logic (Sternberg, 1969), the additive effects of degradation and word frequency have been viewed as suggesting that these two factors influence separate stages in the performance of the lexical decision task.[1] For example, consider the additive effects of degradation and frequency within Becker's (1980) activation–verification model: Word recognition reflects a frequency-ordered serial search of lexical representations that are similar to the target stimulus. Visual degradation primarily influences an early clean-up stage that is needed to identify the features of the stimulus that will be used for the later ordered search process. In this way, stimulus degradation influences relatively early visual clean-up processes, whereas word-frequency presumably influences later serial search processes (see Besner & McCann, 1987, for alternative interpretations).

In this experiment, stimuli were presented either intact or visually degraded by the presence of additional illuminated pixels. Word frequency was factorially crossed with stimulus degradation. If stimulus degradation influences an earlier stage in the information flow, compared with word frequency, and if characteristics of the response after response initiation primarily reflect later stages in the information processing stream, then the results of Experiment 2 should yield (a) additive influences of both word frequency and stimulus degradation on onset latencies; (b) an influence of frequency on response dynamics, after response initiation; and (c) little influence of degradation on response dynamics, after response initiation. However, it is quite possible that the ability to influence the characteristics of the response after response initiation is not limited to relatively late processes. If this is the case, then the results of Experiment 2 should yield influences of both frequency and degradation on response onset and on response dynamics, after initiation. In this way, the results of Experiment 2 should help provide information regarding where in the processing system these response force effects may be occurring.

Method

Participants. Thirty-six participants recruited from the Washington University community took part in this experiment. All participants were native English speakers and naive to the goals of the experiment. They were paid $5 for their participation.

Materials. The target word list was generated from the Kučera and Francis (1967) norms. A total of 80 high-frequency words (counts greater than 100 per million) and 80 low-frequency words (counts less than 5 per million) were selected as targets. High- and low-frequency

---

[1] It is worth noting here that there has been considerable disagreement regarding the separability of processing stages by means of the use of additive-factors logic (see, e.g., McClelland, 1979; Pachella, 1974). We acknowledge that it is indeed difficult to unequivocally dissociate processing stages by means of additive-factors techniques. However, our arguments still hold if stimulus degradation primarily, rather than only, influences earlier processes in the information flow.

words were equated on word length (mean = 4.7 letters). Pronounceable nonwords were generated from an additional 80 high- and 80 low-frequency words selected from Kučera and Francis's norms by replacing one or two letters while keeping the same total number of letters. The nonwords were equated on length with the word targets. In addition, a total of 32 additional high- and low-frequency words were selected for practice–buffer items.

Within each of the two experimental blocks, half of each frequency class were presented degraded and half were presented nondegraded. Across the two test blocks, each participant received 40 trials in each of the four cells that were produced by factorially crossing two levels of frequency and two levels of degradation. Half of the nonword stimuli were presented degraded and half were presented nondegraded. Items were counterbalanced across subjects such that both words and nonwords appeared equally often in degraded and nondegraded presentation formats. Stimulus degradation involved identifying the 8 × 9 matrix of pixels that was used for displaying characters. A random subset of 12 of the 72 (8 × 9) pixels per character display area was then selected for illumination. Some of these characters would overlap with the pixels necessary to display a given character and some would not, and this would vary as a function of character and the output from the random selection routine. The functional effect of this degradation was to simply add a noise mask over the illuminated letter string. As we shall see in the *Results* section, this type of stimulus degradation was highly effective.

Procedure. The procedure was precisely the same as that in Experiment 1, with the following exception. In this experiment, participants made word responses by moving the response handle in one direction and nonword responses by moving the response handle in the opposite direction. Half of the participants made word responses in the rightward direction while the remaining half made word responses in the leftward direction. This change was made for two reasons: First, the results of Experiment 1, in conjunction with the results reported by Abrams and Balota (1991), indicate that the type of response did not appear to modulate the influence of word frequency on response dynamics. Second, with this procedure it is also possible to determine if there is an influence of degradation on response dynamics for nonwords and also address any possible Lexicality × Degradation interaction.

Results

Response latencies and characteristics of the response dynamics were calculated as in Experiment 1. Table 2 displays the means of each of these measures as a function of frequency and stimulus degradation for words. The first set of analyses involved a series of 2 (degraded vs. nondegraded) × 2 (high vs. low frequency) within-subjects ANOVAs on each of the dependent measures for only the word stimuli. As shown in Table 2, there were large effects of word frequency in both onset latencies, $F_1(1, 35) = 137.42$, $MSE = 1{,}311.89$, $p < .001$; $F_2(1, 158) = 56.70$, $MSE = 5{,}708.05$, $p < .001$, and accuracies, $F_1(1, 35) = 62.19$, $MSE = 3.70$, $p < .001$; $F_2(1, 158) = 18.06$, $MSE = 3.30$, $p < .001$. In addition, there was a large effect of stimulus degradation in both onset latencies, $F_1(1, 35) = 327.84$, $MSE = 785.60$, $p < .001$; $F_2(1, 158) = 299.35$, $MSE = 312.00$, $p < .001$, and in accuracies, $F_1(1, 35) = 50.80$, $MSE = 2.76$, $p < .001$; $F_2(1, 158) = 8.41$, $MSE = .01$, $p < .001$. However, the interaction between word frequency and stimulus degradation did not approach significance in either response latencies or accuracies, all Fs $< .30$. Thus, this study accomplished the necessary condition of producing purely additive effects of frequency and degradation on response latency. In fact, the word-frequency effect was remarkably similar for degraded (69 ms in latencies and 7% in accuracy) and nondegraded stimuli (72 ms in latencies and 7% in accuracy). According to additive-factors logic (Sternberg, 1969), these data suggest that degradation and frequency were influencing separable stages in the information flow. Moreover, Schweickert (1985, 1989) has demonstrated that it is extremely difficult for a model that does not assume separable stages to account for simultaneous additive influences of factors on both response latencies and accuracies, as we observed in this study.

Table 2

Mean Response Latencies, Accuracies, and Characteristics of the Response After Response Initiation for Experiment 2 as a Function of Word Frequency and Stimulus Degradation

Dependent variable	Nondegraded		Degraded	
	HF	LF	HF	LF
Response latency (ms)^{a,b}	560.0	632.2	646.0	715.4
% correct^{a,b}	98	91	93	86
Characteristics				
Movement duration (ms)^{a,b}	129.5	131.5	132.0	133.2
End velocity^{b} (cm/s)	111.0	108.5	107.0	107.9
Peak acceleration^{b} (cm/s/s)	172.3	170.2	168.3	166.1
20-ms after onset				
Position (cm)	.386	.366	.368	.384
Velocity^{a} (cm/s)	22.03	21.51	21.86	21.25
Acceleration^{a,b} (cm/s/s)	102.27	96.68	97.55	96.02
50-ms after onset				
Position (cm)	1.52	1.46	1.46	1.47
Velocity^{a,b} (cm/s)	54.73	53.45	52.80	51.99
Acceleration^{b} (cm/s/s)	124.3	124.04	120.2	116.7

Note. HF = high frequency; LF = low frequency.

[a] A reliable frequency effect at least at the $p < .05$ level. [b] A reliable degradation effect at least at the $p < .05$ level.

More important, now consider the results of the analyses of the movement dynamics beyond response initiation. Here, the data are quite consistent in indicating that there were again reliable effects of frequency and, more interestingly, consistent effects of stimulus degradation. First, consider the influence of word frequency. Compared with responses to low-frequency words, responses to high-frequency words had reliably (by subjects) shorter total movement durations, $F_1(1, 35) = 4.44$, $MSE = 21.18$, $p = .04$; $F_2(1, 158) = 2.32$, $MSE = 90.42$, $p = .13$, and faster velocities at the end of the movement, $F_1(1, 35) = 6.15$, $MSE = 30.75$, $p = .02$; $F_2(1, 158) = 6.43$, $MSE = 60.03$, $p = .01$. Turning to the response measures after 20 ms into the movement, compared with responses to low-frequency words, responses to high-frequency words had reliably (by items) higher velocities, $F_1(1, 35) = 3.74$, $MSE = 3.07$, $p < .06$; $F_2(1, 158) = 4.33$, $MSE = 6.75$, $p = .04$, and reliably greater accelerations, $F_1(1, 35) = 5.47$, $MSE = 83.36$, $p = .02$; $F_2(1, 158) = 6.36$, $MSE = 112.30$, $p = .01$. Finally, consider the response measures after 50 ms into the movements. Compared with responses to low-frequency words, responses to high-frequency words had higher velocities, $F_1(1, 35) = 4.16$, $MSE = 9.48$, $p = .04$; $F_2(1, 158) = 5.13$, $MSE = 14.23$, $p = .02$.

Now, consider the effects of stimulus degradation. Compared with degraded stimuli, responses to nondegraded stimuli had (a) shorter total movement durations, $F_1(1, 35) = 12.93$, $MSE = 12.31$, $p < .005$; $F_2(1, 158) = 11.36$, $MSE = 121.23$, $p = .001$; (b) reliably (by subjects) faster velocities at the end of the movement, $F_1(1, 35) = 6.15$, $MSE = 30.75$, $p = .02$; $F_2(1, 158) = 1.025$, $MSE = 76.10$, $p = .32$, and greater peak acceleration during the movement, $F_1(1, 35) = 8.69$, $MSE = 68.26$, $p < .01$; $F_2(1, 158) = 5.88$, $MSE = 168.16$, $p = .02$. After 20 ms into the movement, compared with degraded stimuli, nondegraded stimuli already had responses with reliably greater accelerations, $F_1(1, 35) = 4.34$, $MSE = 70.39$, $p = .04$; $F_2(1, 158) = 6.22$, $MSE = 110.25$, $p = .01$. After 50 ms into the movement, compared with degraded stimuli, responses to nondegraded stimuli had both higher velocity, $F_1(1, 35) = 10.08$, $MSE = 10.27$, $p < .005$; $F_2(1, 158) = 11.87$, $MSE = 18.24$, $p = .001$, and greater acceleration, $F_1(1, 35) = 10.64$, $MSE = 109.92$, $p < .005$; $F_2(1, 158) = 13.06$, $MSE = 172.35$, $p < .001$. Finally, as in the analyses of the onset latencies and accuracies, none of the ANOVAs on the response characteristics produced reliable Frequency × Degradation interactions.

In addition to the analyses addressing degradation and word-frequency effects for words, we also conducted a series of 2 (word, collapsed across high- and low-frequency stimuli, vs. nonword) × 2 (degraded vs. nondegraded) within-subjects ANOVAs to address whether there was a differential influence of degradation on word and nonword stimuli. Table 3 displays these data. First, it should be noted that, as expected, there was a large lexicality effect. Specifically, words produced both faster, $F_1(1, 35) = 170.52$, $MSE = 9{,}319.77$, $p < .001$; $F_2(1, 318) = 240.04$, $MSE = 7{,}104.30$, $p < .001$, and more accurate (by subjects), $F_1(1, 35) = 4.79$, $MSE = 14.22$, $p = .03$; $F_2(1, 318) < 1.00$, $MSE = 21.46$, responses than nonwords. These analyses also yielded large effects of stimulus degradation on both response latencies, $F_1(1, 35) = 356.08$, $MSE = 786.06$, $p < .001$; $F_2(1, 318) = 549.11$, $MSE = 2102.84$, $p < .001$, and accuracies, $F_1(1, 35) = 63.44$, $MSE = 8.40$, $p < .001$; $F_2(1, 318) = 17.41$, $MSE = .03$, $p < .001$. There was no evidence of an interaction between degradation and lexicality in either response latencies, $F_1(1, 35) = 1.38$, $MSE = 329.83$, $p = .24$; $F_2(1, 318) < 1.0$, or accuracies, both Fs < 1.00.

Table 3

Mean Response Latencies, Accuracies, and Characteristics of the Response After Response Initiation for Experiment 2 as a Function of Lexicality (Collapsed Across High- and Low-Frequency Words) and Stimulus Degradation

Dependent variable	Nondegraded		Degraded	
	Word	Nonword	Word	Nonword
Response latency (ms)^{a,b}	596.1	697.3	680.7	789.1
% correct^{a,b}	94	92	89	88
Characteristics				
Movement duration (ms)^{a,b}	130.5	136.4	132.6	139.1
End velocity^{b} (cm/s)	109.7	104.2	107.4	102.8
Peak acceleration^{a,b} (cm/s/s)	171.2	151.6	167.2	149.1
20-ms after onset				
Position (cm)	.376	.407	.376	.388
Velocity (cm/s)	21.77	21.67	21.56	21.06
Acceleration^{a,b} (cm/s/s)	99.45	88.05	96.78	86.71
50-ms after onset				
Position^{b} (cm)	1.49	1.44	1.47	1.40
Velocity^{a,b} (cm/s)	54.09	50.77	52.39	49.20
Acceleration^{a,b} (cm/s/s)	124.2	111.7	118.5	108.3

[a] A reliable lexicality effect at least at the $p < .05$ level. [b] A reliable degradation effect at least at the $p < .05$ level.

Next, consider the effects of lexicality and degradation on the characteristics of the response after response initiation. Again, the effects of degradation were relatively large and quite consistent. As shown in Table 3, compared with degraded stimuli, nondegraded stimuli had responses with shorter total movement durations, $F_1(1, 35) = 17.40$, $MSE = 11.58$, $p < .001$; $F_2(1, 318) = 13.34$, $MSE = 189.15$, $p < .001$, greater (by subjects) end velocities, $F_1(1, 35) = 8.41$, $MSE = 14.15$, $p < .01$; $F_2(1, 318) < 1.00$, and greater peak accelerations, $F_1(1, 35) = 11.23$, $MSE = 34.69$, $p < .005$; $F_2(1, 318) = 9.55$, $MSE = 166.01$, $p = .003$. If one considers aspects of the movement 20 ms beyond response initiation, there was a reliable (by items) influence of degradation on acceleration, $F_1(1, 35) = 3.01$, $MSE = 36.89$, $p = .08$; $F_2(1, 318) = 8.77$, $MSE = 139.85$, $p = .004$. After 50 ms, there was a reliable effect (by items) of degradation on the position of the hand, $F_1(1, 35) = 3.95$, $MSE = .05$, $p = .05$; $F_2(1, 318) = 5.67$, $MSE = .03$, $p = .02$, and highly reliable effects of degradation on both velocity, $F_1(1, 35) = 17.01$, $MSE = 5.64$, $p < .001$; $F_2(1, 318) = 20.02$, $MSE = 22.56$, $p < .001$, and acceleration, $F_1(1, 35) = 20.57$, $MSE = 35.93$, $p < .001$; $F_2(1, 318) = 12.80$, $MSE = 193.27$, $p < .001$. Thus, the results of Experiment 2 indicate very large and consistent effects of stimulus degradation on the dynamics of the response after response initiation.

Turning to the influence of lexicality, there was a reliable effect of lexicality on total movement duration, $F_1(1, 35) = 13.02$, $MSE = 105.61$, $p = .005$; $F_2(1, 318) = 32.31$, $MSE = 192.98$, $p < .001$, on end velocities (by items), $F_1(1, 35) = 1.65$, $MSE = 561.99$, $p > .20$; $F_2(1, 318) = 107.86$, $MSE = 67.79$, $p < .001$, and on peak accelerations, $F_1(1, 35) = 23.97$, $MSE = 531.97$, $p < .001$; $F_2(1, 318) = 341.01$, $MSE = 146.75$, $p < .001$. After 20 ms into the movement, there was an effect of lexicality on position of the hand (by items), $F_1(1, 35) < 1.0$, $MSE = .12$; $F_2(1, 318) = 26.95$, $MSE = .01$, $p < .001$, and acceleration, $F_1(1, 35) = 4.30$, $MSE = 964.84$, $p < .05$; $F_2(1, 318) = 44.83$, $MSE = 131.49$, $p < .001$. After 50 ms into the movement, there was an effect of lexicality on both velocity, $F_1(1, 35) = 4.06$, $MSE = 94.18$, $p < .05$; $F_2(1, 318) = 62.04$, $MSE = 23.52$, $p < .001$, and acceleration, $F_1(1, 35) = 6.59$, $MSE = 700.96$, $p < .05$; $F_2(1, 318) = 60.54$, $MSE = 172.54$, $p < .001$. Thus, responses to words were both initiated more quickly and were more forceful once initiated, compared with responses to nonwords. There was no evidence of an interaction between lexicality and degradation in any of the analyses of the dependent measures that reflected response characteristics.

Discussion

There are a number of noteworthy aspects of Experiment 2. First, the results of the onset latency data clearly yielded additive effects of degradation and word frequency. According to additive-factors logic, this pattern suggests that stimulus degradation was influencing a distinct, and presumably earlier, stage of processing than was word frequency. Hence, this study provided the necessary condition to address the influence of an earlier independent stage of processing on response dynamics after response initiation. Interestingly, under these conditions, both degradation and word frequency consistently influenced not only onset latencies but also response dynamics after response initiation. In this light, it appears that the characteristics of the response after initiation are not limited to manipulations (i.e., frequency) that affect relatively late stages in the processing stream, and hence stages closer to response programming, but also can be sensitive to manipulations (i.e., stimulus degradation) that affect relatively early stages in the processing stream.

Experiment 3

The goal of this experiment was to investigate the influence of word frequency in lexical decision performance in a different response modality, speech. The importance of investigating the influence of word frequency in a speech response is twofold. First, pronunciation performance is the second major response modality in chronometric studies of lexical processing. Hence, it is important to determine whether characteristics of the response after initiation (e.g., duration) in this response modality are also susceptible to being modulated by variables (e.g., word frequency) that traditionally have been viewed as influencing pattern recognition and decision processes. Second, speech is a highly practiced response modality that some researchers (e.g., Fodor, 1966; McNeill, 1970) have argued has a special status in the cognitive architecture. This contrasts with the rather arbitrary button press used in most chronometric studies. It is quite possible that the pronunciation duration of an isolated word is relatively impervious to factors that modulate the speed to initiate the output.

Interestingly, there has been some controversy regarding the influence of word frequency on speech production. For example, in experiments in which participants read lists of high-frequency words or lists of low-frequency words, Wright (1979) provided evidence that the durations of low-frequency words were longer than the durations of high-frequency words. However, Geffen, Stierman, and Tildesley (1979) and Geffen and Luszcz (1983) argued that there were no effects on durations in reading high- and low-frequency lists of words, and that only the pauses between words were affected. Unfortunately, it is relatively difficult to isolate where an effect is having an influence when lists of different classes of stimuli are presented. For example, if there is an effect on durations in such list reading studies, it is possible that the effect is due to participants accessing the next low-frequency word while articulating the current low-frequency word. Balota and Shields (1988) reported some preliminary evidence of a frequency effect on production durations of isolated words in a more traditional naming task. However, although Balota and Shields attempted to equate a number of phonological characteristics across high and low-frequency words (e.g., length in letters, phonemes, onsets, offsets, etc.), some differences will always remain. Specifically, because different phonological sequences occurred in the high- and low-frequency words, it is unclear whether the observed effect was due to frequency or to differences in the phonological characteristics between the two classes of words. Thus, from the available evidence it is unclear whether word frequency can modulate the duration of a speech response.

In Experiment 3, we developed a modified lexical decision task in which participants pronounced the same arbitrary word, *normal,* for both high- and low-frequency words, and made a keypress for nonwords. Because participants produced the same response, *normal,* for both high- and low-frequency words, the phonological characteristics of the output in the response were totally equated across word frequency. Also, we had participants press a key on nonword trials instead of pronouncing a word aloud in order to decrease the likelihood of increased output competition during the responses to the low-frequency words, compared with the high-frequency words (assuming that low-frequency words might be somewhat closer to nonwords on a relevant familiarity dimension). The question of interest is very simple: Will low-frequency words yield longer production durations of the same arbitrary vocal output (*normal*), compared with high-frequency words?

Method

Participants. Thirty undergraduates recruited from the Washington University community served as participants. Each received $5.00 for participating in this experiment. All participants were native English speakers and naive to the goals of the experiment.

Equipment. The experiment was controlled by an Apple IIe microcomputer that was equipped with a Mountain Hardware clock-card to obtain millisecond resolution. A Gerbrands G1341T voicekey was integrated with the computer to detect voice onsets and continuous speech output. The offset of the production was defined as the first 75 ms of continuous silence after the onset of the production.

Materials. The same materials used in Experiment 2 were used in this experiment. The only difference was that there was no manipulation of stimulus degradation.

Procedure. The following sequence occurred on each trial: (a) a letter string was presented at the center of the CRT; (b) either a keypress or the beginning of the pronunciation of the word *normal* triggered the offset of the letter string; (c) if the response was correct (pronouncing "normal" for words, pressing a key for nonwords), the participant pressed a key to begin a 1,500-ms intertrial interval; if the response was incorrect (i.e., the participant triggered the voicekey on a nonword trial or made a keypress on a word trial), a 150-ms tone was presented followed by an ERROR message presented for 750-ms and then a keypress initiated a 1,500-ms intertrial interval.

Participants first received a practice block of 24 trials, which was followed by two test blocks of 164 trials each. Each test block began with four buffer strings, two words and two nonwords. In addition, each test block included 40 high-frequency words, 40 low-frequency words, and 80 nonwords. With the exception of the buffer trials, all trials were randomized within a block anew for each participant.

Participants were instructed to be as fast and as accurate as possible in making their lexical decisions, that is, either pronouncing the word *normal* on word trials or making a keypress on nonword trials. The experimenter remained with the participant during the first 10 trials to ensure that the participant fully understood the instructions. All participants were tested in a sound-deadened anechoic chamber.

Results

Response latencies above or below 2.5 standard deviations from each participant's mean latency, and production durations above or below 2.5 standard deviations from the participant's mean duration were considered outliers. The mean outlier rate was less than 1% and did not vary as a function of frequency ($t < 1.0$).

The results are clear: Participants were both faster and more accurate in the initiation of their responses to high-frequency words (mean response latency = 636 ms; mean accuracy = 94%) compared with low-frequency words (mean response latency = 706 ms; mean accuracy = 87%), $F_1(1, 29) = 279.59$, $MSE = 264.89$, $p < .001$; $F_2(1, 158) = 80.10$, $MSE = 2280.33$, $p < .001$, and $F_1(1, 29) = 63.79$, $MSE = 7.91$, $p < .001$; $F_2(1, 158) = 20.04$, $MSE = 6.29$, $p < .001$, respectively. More important, there was also a reliable difference in production durations. Specifically, the mean production duration for the word *normal* was 470.2 ms on the high-frequency word trials and 472.4 ms on the low-frequency word trials, $F_1(1, 29) = 4.92$, $MSE = 14.77$, $p < .05$; $F_2(1, 158) = 4.11$, $MSE = 68.92$, $p < .05$, respectively. Although the size of this difference appears to be relatively small, there are two points that should be noted: First, one would expect any difference to be relatively small because participants were producing the same response (*normal*) for both high- and low-frequency words. Second, of the 30 participants in this study, 22 participants produced shorter production durations on high-frequency word trials compared with low-frequency word trials, whereas 7 participants produced reversals and there was one tie. This difference was also reliable when using a sign test, $p < .01$.

Discussion

The results of Experiment 3 clearly yielded an influence of word frequency on the production durations of the same speech response (i.e., *normal*). Thus, it appears that processes involved in the lexical processing of high- and low-frequency words can modulate aspects of the duration of the same speech response. However, before further discussing the results of this experiment, there are two aspects of this study that need extending. First, one might ask whether the meaning of the response term, *normal,* might be more consistent with the high-frequency words (semantically or in frequency of occurrence) than with the low-frequency words. In this sense, the naming response is not totally arbitrary. Thus, in Experiment 4, we chose a nonword response term, *berloe,* that participants produced for both high- and low-frequency words. Here the semantics of the nonword, *berloe,* should not vary with respect to the "normality" of the high- and low-frequency words. The second extension in Experiment 4 is that an experimenter coded the participant's responses on-line. This provided an extra control for any mispronunciations.

Experiment 4

Method

Participants. Twenty-four participants were recruited as volunteers from the Washington University Psychology Department participant pool. All participants were native English speakers and naive to the goals of the experiment.

Apparatus, materials, and procedure. All aspects of the present experiment were identical to the previous study with the following two exceptions: (a) Participants produced the nonword *berloe,* instead of *normal* (as in Experiment 3) on word trials, and (b) on each trial the experimenter (instead of the participant) determined whether the voicekey was triggered by the fluent production of the response *berloe* or some other (dysfluent or extraneous) sound triggered the voicekey.

Results

Response latencies above or below 2.5 standard deviations from the mean of the onset latencies and production durations 2.5 standard deviations above or below the mean of the production durations were treated as outliers. As in Experiment 3, the mean outlier rate was less than 1% and did not vary as a function of word frequency ($t < 1.0$).

The results of this experiment were quite consistent with the results of Experiment 3. Participants were both faster to initiate and more accurate in their responses to high-frequency words (mean response latency = 600 ms; mean accuracy = 90%) compared with low-frequency words (mean response latency = 663 ms; mean accuracy = 82%), $F_1(1, 23) = 169.21$, $MSE = 279.99$, $p < .001$; $F_2(1, 158) = 44.57$, $MSE = 2413.29$, $p < .001$, and $F_1(1, 23) = 56.45$, $MSE = 10.17$, $p < .001$; $F_2(1, 158) = 16.28$, $MSE = 7.31$, $p < .001$, respectively. More important, there was again a reliable effect of frequency on production durations. Specifically, the duration of the same nonword response, *berloe,* was reliably shorter for high-frequency words (389.7 ms) than for low-frequency words (392.0 ms), $F_1(1, 23) = 4.82$, $MSE = 13.55$, $p < .05$; $F_2(1, 158) = 3.95$, $MSE = 126.83$, $p < .05$. There are two further points that should be noted about these data. First, the effect sizes of frequency are remarkably consistent with the effect sizes found in Experiment 3. Specifically, the effect of frequency on response latency, accuracy, and production durations in Experiment 3 was 70 ms, 7%, and 2.2 ms, respectively, whereas the effect of frequency on these same three dependent measures in Experiment 4 was 66 ms, 8%, and 2.3 ms, respectively. Clearly, the type of arbitrary speech response (*berloe* vs. *normal*) did not appear to modulate the frequency effect. The consistency in the data is probably in large part a result of the considerable power per participant/cell. Specifically, there were 80 observations per participant/cell, and across the two experiments there were 4,320 observations contributing to the high- and low-frequency cells (2,400 in Experiment 3; 1,920 in Experiment 4). Second, one should note that the error rate was somewhat higher in Experiment 4 (14%), compared with Experiment 3 (9%). This was likely due to the experimenter being more cautious in eliminating any potential dysfluencies. The interesting point here is that even when this more stringent criterion was used, it did not modulate the size of the effect of word frequency on any of the dependent measures.

Discussion

The results of Experiments 3 and 4 clearly indicate that the influence of word frequency beyond the onset of the response in the lexical decision task was not limited to speeded limb movements but also extended to a different response modality, speeded pronunciation. This effect occurred even though participants were producing the same arbitrary response for both high- and low-frequency words. Moreover, this occurred independent of the lexicality of the response (i.e., *normal* vs. *berloe*). Thus, the influence of word frequency was not limited to the relatively arbitrary limb movement used in our previous studies but also extended to the second major dependent measure (speech) used in chronometric studies.

The observation of frequency effects on the duration of the same arbitrary response is quite important because the emphasis in most models of speech production is in the selection of linguistic elements to be produced in speech (see, for example, Dell, 1986; Levelt, 1989). Within these models, once selection occurs, the elements, be they correct or incorrect, are presumably produced in a relatively fluent manner. In this way, there is not a direct relation between variables related to selection and variables related to execution. The previous emphasis on selection processes is in large part due to the reliance on fluent speech errors (e.g., fluently producing *pass pout* instead of *pass out*) as the primary database for speech production models (e.g., see Cutler, 1980; Fromkin, 1973). These are examples in which speakers incorrectly selected elements and conjoined them often in an apparently fluent fashion. Whether such errors are actually fluent is unclear because most data available regarding speech errors have been obtained by means of the coding of naturally occurring speech and in the few studies of experimentally induced speech errors (e.g., Baars, Motley, & McKay, 1975; Dell, 1986), there have not been duration measures available. Our results suggest that factors that presumably influence relatively early stages related to the selection of linguistic elements in speech (e.g., word frequency) can still have influences on the manner in which the same phonological sequence (*normal* or *berloe*) is fluently produced. Obviously, one might question the relevance of the present vocal lexical decision task to processes involved in naturally occurring speech production. In this light, it should be emphasized that the major point that can be made from the naming results is that the speech output system is not impenetrable by variables related to earlier pattern recognition and selection processes.

General Discussion

The primary goal of these experiments was to further extend chronometric investigations into aspects of responses that are beyond the triggering of a microswitch. In this investigation, we have emphasized the role of word frequency in lexical decision performance because this variable has played a central role in the development of word recognition models within the chronometric tradition. In this light, there are a number of noteworthy aspects of the present results. The results of the first experiment yielded a reliable influence of word frequency on response dynamics (after response onset) under conditions that minimized peripheral response competition effects (i.e., moving the same limb in opposing directions for word and nonword responses). In Experiment 2, we manipulated stimulus degradation (along with word frequency) to determine whether response dynamics would be influenced by a variable that had a relatively earlier influence in the information processing stream. The results of this experiment yielded additive effects of word frequency and stimulus degradation on both response latency and accuracy. More important, the results of this experiment also yielded clear effects of both variables on response dynamics after response initiation. Thus, a factor (e.g., stimulus degradation) that has a relatively early influence in the processing stream can also influence the force of a response after its initiation. The results of Experiments 3 and 4 extended the influence of word frequency on the dynamics of limb movements to a completely different response modality, speech. To be specific, these experiments indicated that the duration of the production of the same arbitrary speech response (*normal* in Experiment 3, *berloe* in Experiment 4) was shorter when produced in response to high-frequency words than when produced in response to low-frequency words. These results suggest that speech is no less penetrable than limb movements by a factor (word frequency) that presumably influences pattern recognition and response selection processes. We now turn to a more detailed discussion of the implications of these results.

Insights Into the Information Processing Stream From Response Force Effects

One might ask at this point, what is really being learned about the information processing system above and beyond what we already know from simple response latency measures? In general, it appears that responses that lead to faster onset latencies are also more forceful (also see Angel, 1973; Giray & Ulrich, 1993). Although, as discussed below, this is an overgeneralization, let us consider the implications of this simple observation. If this is all we have learned, we believe that this observation is quite important because such a generalization immediately allows one to reject an implicit assumption in most information processing models. Specifically, it is typically assumed that most variables primarily influence pattern recognition and response selection processes, and once the response has been selected the participant initiates the same ballistic response independent of the influence of the variable in question. To illustrate this assumption, consider a traditional logogen-type (Morton, 1969) account of the effect of word frequency on lexical decision performance. A stimulus word is presented and featural detectors begin to accumulate information at some constant rate independent of frequency. These detectors presumably feed logogens (word recognition devices). The word-frequency effect occurs within this framework because the functional thresholds for the logogens corresponding to high-frequency words are lower than for the logogens corresponding to low-frequency words. Once the threshold is surpassed, the participant simply initiates the word response, and, because of the differences in thresholds, this is a bit later in time for low-frequency words than for high-frequency words. Our results indicate that at least one aspect of this framework is incorrect. Specifically, the word response is not simply ballistically initiated, independent of the factor being manipulated, but the response itself also carries with it information regarding frequency of occurrence. Thus, in this light, the response system is not simply a dedicated module that blindly carries out the commands of a central decision maker. Our results are much more consistent with a cascadic flow of information (McClelland, 1979) in which the influence of factors on early stages is not isolated from systems that directly modulate later response dimensions.

Of course, above and beyond the dismissal of the aforementioned strict serial stage model of binary tasks, there is also evidence that the assumption that response latency and force will always be positively related is an overgeneralization. Consider, for example, the memory scanning results obtained by Abrams and Balota (1991). The results of that study indicated that on positive probe trials (i.e., when the target probe was in the memory set) small memory sets (set size two) led to faster and more forceful responses than did large memory sets (set size six). On negative probe trials (i.e., when the target probe was not in the memory set) small memory sets led to faster onset latencies but, interestingly, less forceful responses than did large memory sets. In accounting for these results, Abrams and Balota presented an extension of Sternberg's (1969) serial exhaustive scanning model in which participants weighed information in support of a yes response against the amount of information in support of a no response. Although on negative probe trials, participants were able to scan and compare two items faster than six items (leading to the obtained difference in response latency), there was more evidence in favor of a no response after six mismatches compared with two mismatches (leading to more forceful responses for the set-size-six trials, compared with the set-size-two trials). The important point for the present discussion is that response latency and response force are not always positively related, and conditions wherein such a relation does not exist may be especially insightful regarding the underlying processes involved in a given task (also, see Giray & Ulrich, 1993).

The results from Experiment 2 also yielded evidence that the simple observation that response latency will always be positively correlated with response force is an overgeneralization. Specifically, the results of Experiment 2 provided evidence that the influence of different factors (frequency and degradation) may be decoupled at different points in time within the same movement. Consider, for example, the influences of frequency and degradation on acceleration after 20 ms into the movement and after 50 ms into the movement. As one can see in Table 2, early in the movement, that is, 20 ms after initiation, there was an effect of frequency on acceleration for the nondegraded stimuli, whereas there was little influence of frequency for the degraded stimuli. Interestingly, however, after 50 ms there was an influence of frequency on acceleration for the degraded stimuli but not for the nondegraded stimuli. In fact, an ANOVA that included time (20 ms vs. 50 ms) as a variable yielded a reliable Frequency × Degradation × Time interaction, $F(1, 35) = 4.57$, $MSE = 52.74$, $p < .05$. Thus, the influence of frequency on force of the movement appeared to be earlier for nondegraded stimuli than for degraded stimuli. Of course, in retrospect, this pattern appears quite reasonable, if one assumes that the operations carried out in perception have carryover effects that retain their temporal ordering into action systems. Specifically, because degraded stimuli need to involve a clean-up process before frequency can exert its influence, the influence of frequency will be later in time for degraded stimuli compared with nondegraded stimuli. Thus, it appears that the temporal dynamics of the movement after initiation can provide a window into the information flow from earlier stages. The important point is that if one only has a single response latency measure in this paradigm, such carryover effects would be impossible to detect.[2]

Converging Evidence of Frequency Effects Beyond Onsets From Typing and Eye-Movement Experiments

It is worth noting here that there is evidence that word frequency can modulate factors beyond response onset in other dependent measures. Consider, for example, skilled typing performance. Here, one might consider the onset of the response the first keypress to initiate the typing of a word. Interestingly, Gentner, Larochelle, and Grudin (1988) reported evidence that there were reliable effects of word frequency on interstroke intervals even under conditions in which (a) the physical difficulty of typing high- and low-frequency words was equated and (b) differences in letter transition frequencies were partialed out. Thus, the Gentner et al. study provides evidence of a frequency effect after the initiation of the response for skilled typists. Likewise, Inhoff (1991) provided evidence that there was a characteristic slowing of the interstroke intervals in the typing of the middle letters in long low-frequency words, but not in long high-frequency words. Inhoff viewed these results as being consistent with a lexical motor programming hypothesis in which lexical-level information plays a role in the development of appropriate motor programs. The important point for the present discussion is that programs for motor output in skilled typing are not impervious to word frequency.

A second domain where one finds influences of frequency beyond the onset of a response is from the eye-tracking literature. In these studies, eye position is monitored as participants are reading text. Rayner and Duffy (1986) have reported evidence that word frequency not only modulated how long participants fixated on a given high- or low-frequency word within the text, but also how long the eyes spent on the next word within the text. Of course, it is possible that this carryover word-frequency effect may have been reflecting postaccess integration processes in reading. The important

---

[2] One might suggest here that the effects observed in onset latencies are much larger than the effects observed in measures of response dynamics. Hence, researchers can be relatively comfortable in ignoring response dynamics in chronometric studies. There are three replies to such an argument. First, although the effects are quite small, the unaccounted-for variance in these measures is also quite small, and hence, these effects are consistently reliable. Second, as noted, the direction of the influence of an independent variable on response dynamics and onset latencies is not always the same, and hence, response dynamic measures can provide valuable insights into earlier processes in the information stream. Third, the size of the effects in no way compromises the important theoretical inference that response systems do not blindly execute responses independent of the variable of interest.

point here is that when one has a continuous record of performance during reading, by means of eye-tracking systems, the influence of word frequency is not restricted to the time spent on the target word, reflecting lexical access and word recognition processes, but also can continue to influence performance after the eyes have left the target word.

Possible Mechanisms Underlying the Force Effects in the Lexical Decision Task

In any chronometric endeavor, a primary goal is to understand the mechanism(s) underlying the variable that is influencing performance. The present study provides some evidence regarding possible mechanism(s) for the obtained force effects in the lexical decision task. As we noted earlier, a number of researchers have suggested (e.g., Balota & Chumbley, 1984; Besner & Swan, 1982; Gernsbacher, 1984; Seidenberg & McClelland, 1989) that one dimension that may be relevant for discriminating words from nonwords in the lexical decision task is the familiarity of the letter string. Words are more familiar patterns than nonwords. Moreover, because high-frequency words should have higher familiarity values than low-frequency words, it is possible that part of the word-frequency effect on response latencies in the lexical decision task may be due to the fact that high-frequency words are more discriminable from nonwords on the relevant familiarity dimension than are low-frequency words.

Given the above analysis of word-frequency effects on onset latencies, there are at least two ways in which the present effects of word frequency on response force measures may have occurred. For simplicity, the first model will be referred to as the confidence model. According to this model, participants may have access to the rate at which information accumulates in the word and nonword response channels and may use the difference between the rate of accumulation of evidence in these two response channels to modulate their confidence in a given response. For high-frequency words, there should be a relatively large difference in the rate of information accrual (high confidence) between the word response channel and the nonword response channel. However, for low-frequency words there will be a relatively small difference in the rate of information accrual (low confidence) in the word response channel and the nonword response channel. This difference between evidence accrual rates for word and nonword response channels (i.e., confidence) can then be used to adjust the force parameter of the response once the initiation threshold has been reached. Very simply, more confident decisions, based on the difference between the rate of information accrual in word and nonword response channels, are converted into more forceful responses. It is also worth noting here that the confidence model handles the effect of degradation. Specifically, the difference in the rate of accumulation of information in the word and nonword response channels (confidence) should be greater for nondegraded stimuli than for degraded stimuli. The important point to note about the confidence model for this discussion is that it suggests that the effects of independent variables on response force occur before the initiation of the response. Hence, the effects of variables on response force reflect preprogrammed parameters of the motor program that are set into action once a decision threshold has been reached.

The second model will be referred to as the enabled response model. According to this model, force effects reflect a more direct influence of independent variables on peripheral output systems. Specifically, consider the possibility that there is a cascadic flow of information into appropriate response channels before (and possibly after) the initiation threshold has been reached. If this were the case, then the increased rate of information accrual for high-frequency words compared with low-frequency words would not produce its influence at the central decision stage that transforms confidence, certainty, or both into specific parameters necessary for the execution of the response, but rather, this increased rate of information accrual would continuously prime appropriate action systems used to execute the response, even before the central decision has been made to initiate the response. This could involve anything from activating appropriate motor areas in the cortex that are involved in making a given hand movement (or particular speech production) to actually changing the tension in appropriate muscle configurations for a given response. Because information should accrue at a faster rate for high-frequency words than for low-frequency words, this would translate into more priming of the appropriate response channels for these items and hence a more fluent response once the response execution threshold has been reached. The enabled response model can also nicely handle the effects of stimulus degradation on response force. Very simply, one would expect more priming of appropriate response channels for stimuli that produce a higher rate of information accrual (i.e., nondegraded stimuli) compared with degraded stimuli.

Although both models can account for the main effects of word frequency and stimulus degradation on limb movements, there are two aspects of the present data that are more consistent with the enabled response model. First, consider the finding that the influence of frequency on response force is earlier within the movement after initiation for nondegraded stimuli than for degraded stimuli. It would be unclear within the confidence model how this pattern would occur. Specifically, within the confidence model, the rate of accumulation within a response channel is translated into force at the time the decision threshold is surpassed, and hence, there is little flexibility in how the response could be modulated during its execution. Alternatively, because the response enabled model involves a continuous activation of the appropriate output systems as the stimulus is processed, it is more plausible that the movement itself will provide a reflection of these earlier processes throughout the movement execution. Very simply, factors that have earlier influences in the system (such as the frequency effect for nondegraded stimuli) are more likely to have earlier influences on the response during the execution of the response because of the additional time available to prime the components involved in the response.

The results of the pronunciation tasks are even more helpful in discriminating between the two models. Consider the notion that confidence is simply translated into force of a given response. Within the limb movement paradigm, the relation between confidence and force would appear to be quite straightforward, that is, more confidence leads to greater force in movement of the limb. However, within the pronunciation paradigm, the translation of confidence to force is a bit less transparent. In fact, within the speech domain, one might assume that greater force might be translated into greater amplitude, as in situations where a speaker places stress on a syllable within speech. If this were the case, then one would expect just the opposite of the obtained effect. Specifically, one might expect responses to high-frequency words to be produced with greater force and hence greater amplitude and duration than responses to low-frequency words.

Although the effect of frequency on durations in pronunciation does not naturally follow from the confidence model, this effect can be accommodated by the enabled response model. The notion here is that the systems necessary for the production of the speech response are receiving partial activation before the response is initiated (i.e., the appropriate action systems have been primed more for high-frequency words than for low-frequency words). Hence, once the response threshold has been reached, the correct response may be produced in a more fluent manner. To be more specific, consider the possibility that when the execution threshold has been reached, there are *N* possible relevant dimensions, most of which are consistent with the target response and probably some of which are inconsistent with the target response. The fluency of the execution of a response reflects at some level the difference between the activation levels of compatible response dimensions and the activation levels of incompatible response dimensions. The enabled response model simply suggests that the dimensions compatible with the correct response have received some priming, or the dimensions that are incompatible with the correct response have received some inhibition, or both, before the response threshold has been reached. The important point here is that priming of motor systems does not necessarily reflect a change in the force of an action, but such priming should produce responses that are more fluent. Of course, the fluency of a limb movement may produce a response that is more forceful. However, the fluency of a speech response does not necessarily translate into a more forceful response. In fact, fluency within this system is more likely to produce a response that is shorter and clearer, that is, closer to the ideal vocal target tract configuration (see MacNeilage, 1970). In this way, the enabled response model can account for both the influence of frequency on the apparent force of the limb movements and the modulation of the duration of the pronunciation response.

In addition to the present pronunciation data, it is noteworthy that there are data by Coles et al. (1985) that also appear more consistent with the enabled response model. As noted earlier, Coles et al. used a simple flanker task in which participants were required to squeeze a dynamometer with the left hand if a given target letter (e.g., *H*) was presented foveally and make a squeeze response with the right hand if a different letter (e.g., *S*) was presented foveally. Target foveal stimuli were also presented with flanking stimuli. On some trials, the flanking stimuli were compatible with the foveally presented target letter (e.g., *SSSSS* or *HHHHH*), and on other trials the flanking stimuli were incompatible with the foveally presented target letter (e.g., *SSHSS* or *HHSHH*). One of the interesting findings in this study was that on some trials, electromyographic activity was present in the forearm of the incorrect limb under the incompatible flanking conditions, even though participants eventually produced the correct response. This would appear to suggest that there was continuous flow into more peripheral output systems (muscle tension) even on trials in which such responses were not eventually initiated (see also DeJong, Coles, Logan, & Gratton, 1990). Such continuous flow within the present context could be viewed as enabling the appropriate peripheral output systems so that the response is executed more forcefully once the initiation threshold has been surpassed. It is unclear how such activation of incorrect response systems on correct response trials could be handled within the confidence–force model.

Conclusions

Our results provide clear evidence that the effects of word frequency extend into response dynamics in the lexical decision task, beyond response initiation. This influence appears to occur independent of type of manual response (i.e., either opposing or nonopposing limb movements) or modality of response (i.e., limb movement or speech). In addition, we have demonstrated that a factor that presumably influences a relatively early and distinct stage in the information flow, stimulus degradation, can also influence the dynamics of the response after response initiation. Hence, the present response force effects do not appear to be limited to the influence of frequency in the lexical decision task. Although alternative models are possible, we feel that these results are most consistent with an enabled response framework in which early operations begin to enable more peripheral systems necessary to implement a given response.

In summary, we feel that along with psychophysiological measures (e.g., event related potentials) that allow one to decouple relatively early analyses before the response onset, careful analyses of output after response onset will help to provide a more complete model of the information processing stream (see also Giray & Ulrich, 1993; Ulrich & Wing, 1991). Although not an arbitrary moment, the triggering of a microswitch is only a single moment in the cascade of events that occur between perception and action. There is no a priori reason for our investigations to be limited to this particular moment.

References

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2,* 153–157.

Angel, A. (1973). Input–output relations in simple reaction time experiments. *Quarterly Journal of Experimental Psychology, 25,* 193–200.

Baars, B. J., Motley, M. T., & McKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14,* 382–391.

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

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

Balota, D. A., & Chumbley, J. I. (1990). Where are the effects of frequency in visual word recognition tasks? Right where we said they were! Comment on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General, 119,* 231–237.

Balota, D. A., & Shields, L. (1988, November). Localizing word-frequency effects in pronunciation. Paper presented at the 29th Annual Meeting of the Psychonomic Society, Chicago.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8,* 493–512.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 389–401.

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and production: An examination of four classes of models. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 201–219). Hillsdale, NJ: Erlbaum.

Besner, D., & Swan, M. (1982). Models of lexical access in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33(A),* 313–325.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840.

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 529–553.

Cutler, A. (1980). Errors of stress and intonation. In V. Fromkin (Ed.), *Errors of linguistic performance: Slips of the tongue, ear, pen, and hand* (pp. 67–80). New York: Academic Press.

DeJong, R., Coles, M. G. H., Logan, G. D., & Gratton, G. (1990). In search of the point of no return: The control of response processes. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 164–182.

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

Fodor, J. A. (1966). How to learn to talk: Some simple ways. In F. Smith & G. A. Miller (Eds.), *The genesis of language: A psycholinguistic approach* (pp. 105–123). Cambridge, MA: MIT Press.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31,* 307–319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26,* 489–504.

Fromkin, V. A. (1973). Slips of the tongue. *Scientific American, 229,* 110–116.

Geffen, G., & Luszcz, M. A. (1983). Are the spoken durations of rare words longer than those of common words? *Memory & Cognition, 11,* 13–15.

Geffen, G., Stierman, I., & Tildesley, P. (1979). The effect of word length and frequency on articulation and pausing during delayed auditory feedback. *Language and Speech, 22,* 191–199.

Gentner, D. R., Larochelle, S., & Grudin, J. (1988). Lexical, sublexical, and peripheral effects in skilled typewriting. *Cognitive Psychology, 20,* 524–548.

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

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1278–1291.

Inhoff, A. W. (1991). Word frequency during copytyping. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 478–487.

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

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197,* 792–795.

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

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6,* 172–187.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review, 77,* 182–196.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McNeill, D. (1970). *The acquisition of language: The study of developmental psycholinguistics.* New York: Harper and Row.

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review, 95,* 340–370.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26,* 3–67.

Monsell, S. (1990). Frequency effects in lexical tasks: Reply to Balota and Chumbley. *Journal of Experimental Psychology: General, 119,* 335–339.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43–71.

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

Norris, D. G. (1984). The effects of frequency, repetition, and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36(A),* 507–518.

Osman, A. M., Kornblum, S., & Meyer, D. E. (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 243–258.

Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, NJ: Erlbaum.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201.

Schuberth, R. E., Spoehr, K. T., & Lane, D. M. (1981). Effects of stimulus and contextual information on the lexical decision process. *Memory & Cognition, 9,* 68–77.

Schweickert, R. (1985). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin, 97,* 530–546.

Schweickert, R. (1989). Separable effects of factors on activation functions in discrete and continuous models: d' and evoked potentials. *Psychological Bulletin, 106,* 318–328.

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

Shields, L. W., & Balota, D. A. (1991). Repetition and associative context effects in speech production. *Language and Speech, 34,* 47–55.

Stanners, R. F., Jastrzembski, J. E., & Westbrook, A. (1975). Frequency and visual quality in a word–nonword classification task. *Journal of Verbal Learning and Verbal Behavior, 14,* 259–264.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276–315). Amsterdam: North-Holland.

Ulrich, R., & Wing, A. M. (1991). A recruitment theory of force–time relations in the production of brief force pulses: The parallel force unit model. *Psychological Review, 98,* 268–294.

Wilding, J. M. (1988). The interaction of word-frequency and stimulus quality in the lexical decision task—Now you see it, now you don't. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38(A),* 213–228.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition, 7,* 411–419.

Received June 10, 1993
Revision received October 6, 1994
Accepted November 18, 1994 ■
