Journal of Experimental Psychology:
Human Perception and Performance
2005, Vol. 31, No. 5, 1083–1095

Copyright 2005 by the American Psychological Association
0096-1523/05/$12.00 DOI: 10.1037/0096-1523.31.5.1083

Characterizing the Motor Execution Stage of Speech Production: Consonantal Effects on Delayed Naming Latency and Onset Duration

Kathleen Rastle
Royal Holloway, University of London,
and Macquarie University

Karen P. Croot
University of Sydney and Macquarie University

Jonathan M. Harrington
University of Kiel and Macquarie University

Max Coltheart
Macquarie University

The research described in this article had 2 aims: to permit greater precision in the conduct of naming experiments and to contribute to a characterization of the motor execution stage of speech production. The authors report an exhaustive inventory of consonantal and postconsonantal influences on delayed naming latency and onset acoustic duration, derived from a hand-labeled corpus of single-syllable consonant–vowel utterances. Five talkers produced 6 repetitions each of a set of 168 prepared monosyllables, a set that comprised each of the consonantal onsets of English in 3 vowel contexts (e.g., /sli/, /sla/, /slɔː/). Strong and significant effects associated with phonetic characteristics of initial and noninitial phonemes were observed on both delayed naming latency and onset acoustic duration. Results are discussed in terms of the biomechanical properties of the articulatory system that may give rise to these effects and in terms of their methodological implications for naming experiments.

Keywords: articulation, naming latency, reading aloud, initial phoneme, speech production

The production of speech can be conceptualized as involving a number of processing stages including the generation of an abstract phonological representation from print or conceptual knowledge, the computation of a plan or program for the articulatory realization of that phonological representation, and the motor execution of that plan or program (Levelt, 1989). Each of these broad levels of processing can be further subdivided: For example, the generation of an abstract phonological representation from the printed word may occur through dictionary lookup or by rule (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), the articulatory plan may be retrieved from a syllabary or assembled on the fly (Levelt & Wheeldon, 1994), and the motor execution of that plan may be initiated only once that plan has been retrieved and unpacked (Monsell, 1986; Sternberg, Knoll, Monsell, & Wright, 1988; Sternberg, Monsell, Knoll, & Wright, 1978).

This article contributes to the development of a temporal characterization of the motor execution stage of processing. Specifically, we examine the temporal relationship between three events that occur during this stage of processing: the onset of motor execution, the onset of acoustic energy, and the acoustic onset of the vowel. The onset of motor execution is here conceptualized not as an observable physical event (e.g., the first movement of the lips or the first detectable activity of the laryngeal muscles) but as the moment at which the cognitive plan for speech is delivered to the speech execution system, initiating coordinated movement of the articulators. The onset of acoustic energy is a consequence of air moving across already active vocal organs (e.g., the lungs, larynx, and supralaryngeal articulators) and thus always occurs some milliseconds after the onset of motor execution. Likewise, when consonant–vowel (CV) syllables are considered, the acoustic onset of the vowel always occurs some milliseconds after the onset of acoustic energy. For example, at the moment at which the cognitive plan for the syllable /pa/ is passed to the speech execution system (i.e., at the onset of motor execution), a number of articulatory events begin, including the buildup of air pressure behind an airtight closure formed at the lips. The onset of acoustic energy occurs only on the burst release of this airtight closure, some milliseconds after the onset of articulation (Halle, Hughes, & Radley, 1957). Finally, we define the acoustic onset of the vowel in the example /pa/ to occur at voicing onset some milliseconds later (i.e., when the vocal folds begin to vibrate, resulting in an acoustically periodic waveform).

In this article, we investigate two time intervals derived from these three execution-level events, occurring as monosyllables are being uttered. First, we investigate the time taken to produce

---

Kathleen Rastle, Department of Psychology, Royal Holloway, University of London, Egham, Surrey, United Kingdom, and Speech, Hearing, and Language Research Centre, Macquarie University, Sydney, New South Wales, Australia; Karen P. Croot, Department of Psychology, University of Sydney, Sydney, New South Wales, Australia, and Speech, Hearing, and Language Research Centre, Macquarie University; Jonathan M. Harrington, Institute of Phonetics and Digital Speech Processing, University of Kiel, Kiel, Germany, and Speech, Hearing, and Language Research Centre, Macquarie University; Max Coltheart, Macquarie Centre for Cognitive Science, Macquarie University.

This research was supported by the Australian Research Council. We are indebted to Marc Brysbaert, John Clark, William Connolly, Matt Davis, and Sallyanne Palethorpe for helpful discussion; to Zoe Evans for research assistance; and to Chris Callaghan for technical expertise.

Correspondence concerning this article should be addressed to Kathleen Rastle, Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom. E-mail: Kathy.Rastle@rhul.ac.uk

acoustic energy once motor execution has been initiated. We refer to this interval as the execution–acoustic interval (EAI). Second, we investigate the acoustic duration of the syllable onset (OD)—the time elapsing between the acoustic onset of the monosyllable and the acoustic onset of the vowel. We are concerned with both (a) the extent to which these two intervals differ across the consonant phonemes of English (when these consonant phonemes appear in singleton and cluster contexts and when they appear in a range of vowel contexts) and (b) the extent to which these differences can be understood as a function of the articulatory properties of English consonants. Our analyses of EAI and OD differences across consonants address both methodological and theoretical issues pertaining to the study of speech production.

Methodological Significance

From a methodological perspective, it is important to quantify execution-level differences in EAI and OD across consonant phonemes so that measures of speech behavior can be related to cognitive models of speech production. Although cognitive models of speech production have typically been concerned with processes that occur prior to the onset of motor execution, the behavioral measures that inform these theories—most significantly, naming latency and, more recently, production duration (Balota & Abrams, 1995; Balota, Boland, & Shields, 1989; Kawamoto & Kello, 1999; Kawamoto, Kello, Jones, & Bame, 1998; Kello & Plaut, 2000; Kello, Plaut, & MacWhinney, 2000)—reflect events that occur subsequent to the onset of motor execution. Measures of word or picture naming latency, for example, include the EAI, and measures of production duration include OD. If there were differences across phonemes in these intervals, then it would be important to control them—lest any differences arising at the execution level of processing erroneously be attributed to cognitive processes that precede the onset of execution. Since the late 1980s, many naming researchers have attempted to deal with this potential problem by controlling the phonetic characteristics of stimuli across experimental conditions. For example, researchers have controlled the initial phoneme (e.g., Taraban & McClelland, 1987), the phonetic class of the initial phoneme (e.g., Coltheart & Rastle, 1994), or the complete syllabic onset (i.e., all phonemes before the vowel; e.g., Kawamoto et al., 1998)—though these types of matching conventions do not yet appear to be obligatory in naming research (e.g., Monaghan & Ellis, 2002). Indeed, because execution-level effects have not been well quantified, it is not yet known which (if any) phonetic characteristics need to be controlled in psychological studies measuring speech behavior or what the implications are of failing to achieve this control.

Theoretical Significance

From a theoretical perspective, speech motor execution is a domain of human performance that is not yet well understood. A great deal of research in the discipline of experimental phonetics has, of course, been devoted to developing a specification of the biomechanical aspects of speech production that yield the acoustic signal. Progress in this area has been achieved through the use of multiple approaches including sophisticated acoustic (e.g., Lisker & Abramson, 1964), instrumental (e.g., Gracco & Löfqvist, 1994), articulatory imaging (e.g., Narayanan, Alwan, & Haker, 1995), and modeling (e.g., Stevens, 1971) techniques. Although measures of acoustic duration (e.g., OD) have been central to this research stream, rarely have measures of the EAI been used as an investigative tool. We believe, however, that EAI differences across consonants can inform our understanding of speech motor execution—and, conversely, that any characterization of speech motor execution should be able to account for these differences. Theoretical cognitive psychology offers a special contribution to this largely phonetic domain because it provides the means to isolate the motor execution level of processing. Our aim is therefore not simply to describe the EAI and OD differences across consonants but to consider our data in light of what is known about speech motor execution with the goal of understanding why these differences occur. In general, we believe that where time can provide an index of "mental work" when a cognitive system is being studied (from which models of mental architectures and processing components can be drawn), time also indexes the mechanical activities occurring within the physical system responsible for articulation (and studies of these intervals may thus contribute similarly to detailed models of this physical system).

Isolating the Speech Motor Execution Stage of Processing

Speech production tasks typically used in the psychological literature—for example, picture naming, reading aloud, and repetition—reflect a number of cognitive processes, including perceptual processes, conceptual processes, phonological retrieval processes, motor planning processes, and execution processes (Levelt, 1989). We wish to quantify consonantal differences in EAI and OD and so require a task that eliminates, as much as possible, processes that occur prior to the onset of motor execution. For this reason, we used a delayed naming task, in which participants produced already prepared syllables as quickly as possible after a signal to respond. For each syllable, measures of OD were calculated as the interval between the acoustic onset of the word and the acoustic onset of the vowel (with these acoustic boundaries determined according to the criteria specified in the Method section). Measures of EAI were less straightforward because the onset of motor execution does not correspond to an observable event. We instead measured the interval between the signal to respond and the onset of acoustic energy—an overestimate of the EAI. In this section, however, we argue that any variation in this measurement across consonantal onsets must be due to variation in the EAI itself, as long as particular conditions are met.

This argument is based on the theory of the control and execution of prepared action sequences developed by Sternberg, Monsell, and their colleagues (Monsell, 1986; Sternberg et al., 1978, 1988). These authors conducted numerous experiments requiring skilled participants to fully prepare a speech "unit" or series of speech "units," to be produced as a single fluent utterance following a signal to respond. Their findings were remarkably clear. The number of units in the prepared list had a positive and approximately equivalent influence on both (a) the time interval between the signal to respond and the acoustic onset of the first unit and (b) the mean unit duration (acoustic offset of the unit – acoustic onset of the unit). Each of these dependent variables (latency and unit duration) was also influenced independently by the size of the units in the list (e.g., number of syllables). This body of observations formed the basis for their theory of responding in this task (illustrated in Figure 1), according to which a fully specified motor plan is maintained in a buffer until the signal to respond occurs,

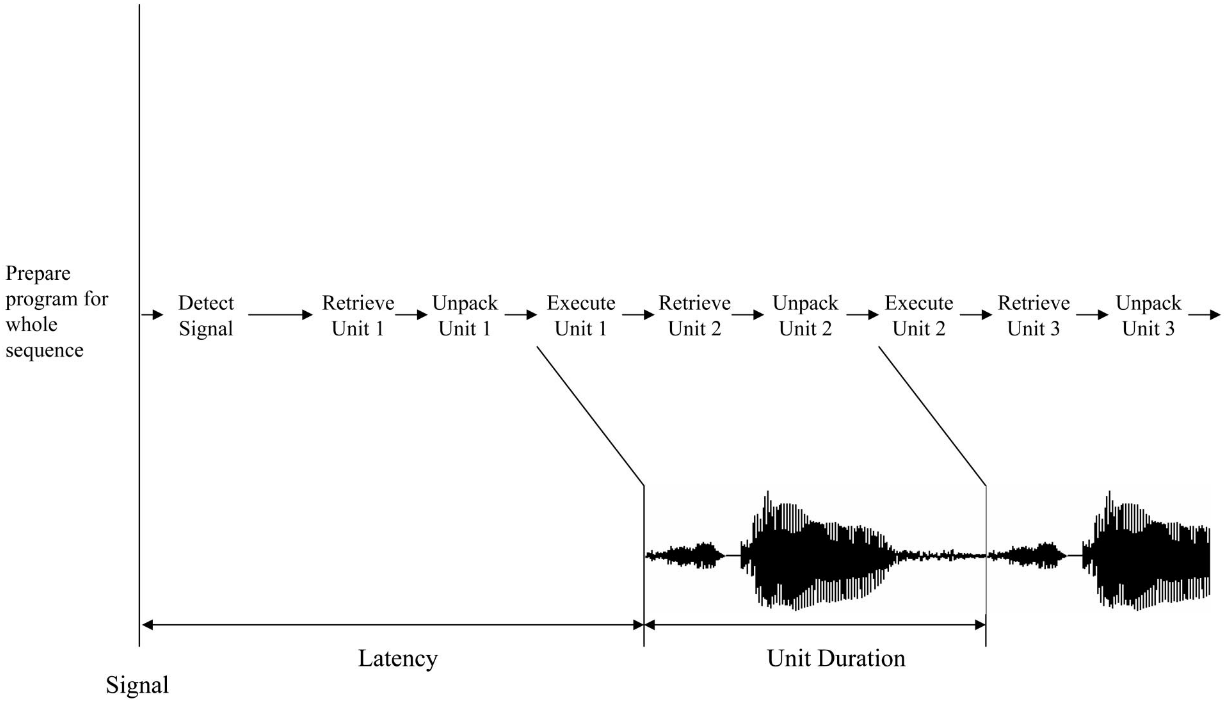

Figure 1. A model of the control and execution of prepared sequences of utterances developed by Sternberg et al. (1978). From "Programming of Complex Sequences: Evidence From the Timing of Rapid Speech and Other Productions," by S. Monsell, in *Generation and Modulation of Action Patterns* (p. 77), edited by H. Heuer and C. Fromm, 1986, Heidelberg, Germany: Springer-Verlag. Copyright 1986 by Springer-Verlag. Adapted with permission.

after which its contents are retrieved, unpacked into their constituents (e.g., syllables, articulatory gestures), and executed.

As is clear from Figure 1, response latencies comprise the sum of (a) the time taken to detect the signal to respond, (b) the time taken by one retrieval stage, and (c) the time taken by one unpacking stage. Unit durations are the sum of (a) one execution stage, (b) one retrieval stage (for the next unit in the program), and (c) one unpacking stage (for the next unit in the program). According to the theory, retrieval time is a function of the number of units in the plan, and unpacking time is a function of the size of each unit. A unit, according to the theory, is one stress group. A stress group, according to Sternberg et al. (1988), comprises any utterance that has a single stressed syllable (e.g., both "one" and "one-and-a" comprise one stress group, both "one-two-three" and "one-and-a-two-and-a-three-and-a" comprise three stress groups). The size of a unit is the number of syllables it contains (e.g., although "one-two-three" and "one-and-a-two-and-a-three-and-a" each contain three stress groups, those stress groups are larger in the latter case). The elegance of the theory derives from its ability to account for both the latency and the duration observations within a single set of assumptions.

We have already pointed out that the onset of motor execution does not correspond to an observable event, and so it is not possible to obtain a direct measurement of the EAI. One can, however, measure the interval between the signal to respond and the acoustic onset. As is clear from Figure 1, this interval includes not only the EAI but also three other periods that precede the EAI: the time taken to detect the signal, followed by the time taken to retrieve the plan, followed by the time taken to unpack the plan. Consider how each of these three periods might be affected by the nature of the initial phoneme of the monosyllable to be uttered. Obviously, the time taken to detect the signal will not vary as a function of the initial phoneme but nor, according to the theory described above (Monsell, 1986; Sternberg et al., 1978), will there be any influence of initial phoneme on the duration of either of the next two periods (the time taken to retrieve the plan and the time taken to unpack the plan), provided what is to be uttered is a monosyllable. This is because (a) the time taken to retrieve the plan depends only on the number of stress groups in the utterance, and every monosyllable has exactly one stress group, and (b) the time taken to unpack the plan depends only on the number of syllables in the utterance, and every monosyllable has exactly one syllable. Given these arguments, any variation in the interval between signal onset and acoustic onset due to the nature of the initial phoneme can be ascribed to an influence of that phoneme on the EAI.

An important assumption from Sternberg et al.'s (1978) and Monsell's (1986) theory that we are adopting is that the delayed naming task reflects only those processes that occur subsequent to the compilation of a motor plan for speech (which is held in a buffer until the signal to respond; it is from that buffer that the retrievals referred to in Figure 1 are made). This assumption has, in the past, been contentious largely because of Balota and Chumbley's (1985) observation that delayed naming latency is influenced by word frequency, a variable usually attributed to lexical stages of processing. Monsell (1990; see also Monsell, Doyle, &

Haggard, 1989) argued, however, that such frequency effects are not normally observed in the delayed naming task (citing McRae, Jared, & Seidenberg, 1990; Monsell et al., 1989; Savage, Bradley, & Forster, 1989) and indeed would be expected to occur only in situations in which participants are not adequately prepared to respond. He commented not only on the relatively short delays used by Balota and Chumbley (1985) but also on the unusually long latencies observed in their study—implying that processing was not restricted to postplanning stages. More recent research has supported Monsell's (1990) arguments: Goldinger, Azuma, Abramson, and Jain (1997) reported that word frequency effects appear in the delayed naming task only when delays are so short (e.g., 400 ms or less) as to preclude adequate preparation. We therefore felt confident in accepting the assumption that the delayed naming task reflects only postplanning articulatory operations, as long as it is carefully executed. As described in the Method section, we made some considerable effort to ensure that participants in our study had adequate preparation time before being prompted to respond.

Previous Studies of EAI and OD Differences Across Consonants

According to Monsell (1986), the onset of the execution stage of processing "initiates downstream activity resulting ultimately in a pattern of vocal tract activity appropriate to the speech element specified" (pp. 76–77). Despite the extensive research carried out by Monsell, Sternberg, and their colleagues (Monsell, 1986; Sternberg et al., 1978) using the delayed naming task, however, their aims were not to investigate the specific events occurring at the execution stage of processing, such as the relationship between the onset of motor execution, acoustic onset, and acoustic vowel onset for different consonant phonemes. As a result, Figure 1 (adapted from Monsell, 1986) is underspecified. It equates, for example, the onset of the execution of a unit with the onset of acoustic energy for that unit. This characterization cannot be correct for any segment but is best illustrated by the voiceless oral stops, in which the articulators completely occlude the vocal tract for tens of milliseconds prior to the onset of acoustic energy at the burst release.

Empirical material relevant to the EAI and OD differences under scrutiny here is available, however. In reviewing this literature, we sought empirical material that met two criteria. First, we sought material that isolated, as closely as possible, execution levels of speech processing—although we are unaware of any study that has achieved this deliberately through the principled task analysis that we provide here. Second, we sought material that identified acoustic boundaries not by means of voice key (commonly used in experimental psychology for measurements of naming latency and duration) but rather through visual inspection of the speech waveform or spectrogram. Voice keys are notoriously inaccurate (Rastle & Davis, 2002); further, their error is not constant over all manner classes of phonemes (Sakuma, Fushimi, & Tatsumi, 1997). By excluding empirical material derived from voice key measures (in particular, the large regression studies of word naming; e.g., Kessler, Treiman, & Mullennix, 2002; Spieler & Balota, 1997), we ensured that any EAI and OD differences across consonants observed in previous studies were not contaminated by differences across consonants in voice key detection accuracy.

We are not aware of any research that has sought to quantify EAI differences across consonant phonemes by using a task specifically designed to isolate the execution stage of processing. In fact, we know of only two studies (Fowler, 1979, and Sakuma et al., 1997, who published a study in Japanese) that have even considered these differences and why they might arise—although neither study reported an exhaustive description of EAI differences across consonants or investigated consonantal differences across a range of vowel and consonantal cluster types. These studies focused primarily on the influence of manner class of initial phoneme (e.g., stop, nasal, fricative) on speeded reading aloud latency. There are very good reasons to suspect that such influences should be found—reasons illustrated most clearly by the case of oral stops. Unlike other manner classes of consonants (e.g., fricative, nasal) in which acoustic energy is present during the articulatory closure, the articulation of oral stops (and affricates) is characterized by a complete occlusion of the vocal tract during which time no acoustic energy is present. We would therefore expect to observe longer reading aloud latencies for stops and affricates than for other manner classes of consonants. This finding is exactly what Fowler (1979) and Sakuma et al. (1997) reported. Fowler further reported that latencies were longer for syllables beginning with the voiced stops /b d g/ than for those beginning with the voiceless stops /p t k/. She reconciled this finding with evidence that the articulatory approach to closure is often more rapid for the voiceless stops than for the voiced stops (see MacNeilage & Ladefoged, 1976). Although Fowler used a speeded reading aloud task (thus leaving open the possibility that her effects could be attributed to any level of processing involved in that task), she attributed the effects observed to the different points relative to action initiation at which different consonants have acoustic consequences. Our study goes well beyond these initial investigations by considering numerous other ways (besides, e.g., manner class) in which the articulatory–acoustic relationship differs across different consonants.

Consonant acoustic duration has been much more extensively researched, although again we are unaware of any research that has compiled an exhaustive inventory of OD effects by using a task that isolates execution-level processes. The available studies of consonant acoustic duration have rarely been restricted to the OD interval (e.g., they include analyses of final consonants; e.g., Umeda, 1977), and the tasks used range from prepared production of citation-form speech (e.g., Baum & Blumstein, 1987), to production in a sentence context (e.g., Baum, 1996; Klatt, 1975), to reading aloud extended passages of text (e.g., Umeda, 1977). Nevertheless, a number of interesting generalizations can be derived from the literature, particularly with regard to the oral stops and fricatives. For both oral stops (e.g., Baum & Ryan, 1993; Klatt, 1975; Lisker & Abramson, 1964) and fricatives (e.g., Baum, 1996; Baum & Blumstein, 1987; Umeda, 1977), voiced exemplars are typically shorter in acoustic duration than are voiceless exemplars. For the oral stops, there is a place-of-articulation effect on acoustic duration, such that bilabials are shorter than alveolars, and alveolars are shorter than velars (e.g., Harrington & Cassidy, 1999; Lehiste & Peterson, 1961; Lisker & Abramson, 1964). Similar place-of-articulation differences can also be found in the fricatives, with the alveolars /s z/ longer in acoustic duration than the interdentals /θ ð/ and labiodentals /f v/ (e.g., Baum, 1996; Behrens & Blumstein, 1988; Umeda, 1977). Consonant acoustic duration is also influenced by the characteristics of subsequent segments.

Irrespective of voicing, oral stops preceding high vowels are typically longer in acoustic duration than are those preceding other types of vowels (Klatt, 1975), and the same is true for fricatives (Schwartz, 1969). Further, oral stops and fricatives appearing in clusters are typically shorter in acoustic duration than are those appearing in singleton contexts (e.g., Klatt, 1974; Lindblom & Rapp, 1973; Schwartz, 1970; Umeda, 1977). Our study considers each of these influences on OD as well as potential reasons for why they might occur, within a single controlled environment designed to isolate the motor execution stage of processing.

The Study

In the research presented here, we investigated EAI and OD periods when these periods occurred in productions of prepared monosyllables. The monosyllables had a CV structure and comprised all possible consonantal onsets of English in a range of vowel contexts. All acoustic boundaries were hand marked to ensure that EAI and OD data would not be contaminated by voice key error.

We sought to meet our methodological aims by producing an Appendix and summary table of EAI and OD values across the consonantal onsets and vowel contexts, which can be used as an aid to stimulus selection or in the analysis of naming data. We sought to meet our theoretical aims by seeking to uncover the articulatory parameters (e.g., manner of articulation, voicing, place of articulation) that account for the EAI and OD differences observed.

Method

Participants

Five talkers from Macquarie University, all with a minimum of 5 years of phonetic training (range is approximately 5–25 years), participated in the experiment. All talkers were native speakers of Australian English, had normal or corrected-to-normal vision, and were free of any known speech or hearing deficits. Each talker took part on a voluntary basis and was required to attend two testing sessions separated by at least 1 day.

Stimuli

Stimuli were 168 CV structured syllables comprising 56 consonantal onsets in three vowel contexts (see the Appendix). These onsets were chosen because they comprise the full set of phonotactically legal consonantal onsets of English (see Rastle, Harrington, & Coltheart, 2002), when onsets consisting of /j/ as the second or third constituent (e.g., “cute,” “spew”) are excluded (these onsets are restricted in Australian English to syllables with a /u/ nucleus). The three vowels studied included /i/ (e.g., “heed”); /a/ (e.g., “hard”); and a neutral, midcentral vowel denoted in this article by the symbol /əː/. This neutral vowel was identical to the vowel that would be produced if the words “the” or “a” were uttered in isolation. These vowels were chosen because they vary on two key dimensions relevant to tongue placement: height and backness. These dimensions are reflected in the first two formants (F1 and F2) of a vowel’s acoustic realization, in which height is inversely related to F1 and backness is inversely related to F2. While the vowel /i/ (as in “heed”) is characterized as a high-front vowel in Australian English (mean F1 = 290 Hz, mean F2 = 2250 Hz), tongue placement in the vowel /a/ (as in “hard”) is lower and further retracted in the oral cavity (mean F1 = 740 Hz, mean F2 = 1410 Hz). Tongue position in the production of the neutral vowel /əː/ in a citation-form context is more central on both dimensions, and this is reflected by formant values between those of /i/ and /a/ (mean F1 = 480 Hz, mean F2 = 1500 Hz). These mean formant values are taken from a classic study of vowel production in Australian adult male speakers (Bernard, 1967).

Apparatus and Procedure

Participants were tested individually in a sound-treated room in the Speech, Hearing, and Language Research Centre at Macquarie University. They were seated approximately 40 cm from a 14-in. (35.56-cm) monochrome monitor, on which stimulus materials were presented. Stimulus presentation was controlled by the QPROMPT display system (developed at the Speech, Hearing, and Language Research Centre) running on an Intel 286 personal computer. The computer was connected both to a two-button response box and to an auditory stimulator that produced a 50-ms tone as required. Acoustic recordings were accomplished with a Sennheiser ME80 directional microphone, placed approximately 20 cm from each participant’s mouth.

Stimuli were presented one at a time on the monitor. Each stimulus appeared as a string of phonetic symbols to ensure that the desired pronunciation would be elicited (all talkers had many years of training in pronouncing phonetic transcriptions). Talkers were instructed to practice pronouncing each stimulus aloud several times before assuming an articulatory rest position. The rest position was intended to minimize the possibility that components of the EAI (e.g., the buildup of air pressure behind a complete closure) could be achieved prior to the signal to respond. Participants indicated their preparedness for the speeded trial with a button press. Following a variable delay, a 1000-Hz tone produced by the auditory stimulator sounded for 50 ms. The delay between button press and tone comprised a random value between 1200 ms and 1800 ms, which changed on every trial. A variable delay was used to discourage talkers from anticipating the signal to respond. Talkers were instructed to produce the target syllable as soon as possible after the tone. In summary, the response–stimulus interval (for which the stimulus was the 1000-Hz tone) comprised the time taken by producing the response, requesting the next stimulus, practicing the stimulus aloud as required, assuming the rest position, pressing the button to indicate preparedness, the variable delay, and any breaks that the participant chose to take between trials. Though it varied considerably across trials, the response–stimulus interval averaged around 15 s. We allowed talkers to control the speed of stimulus presentation, both to ensure that they had adequate preparation time for the production of each syllable and to minimize the contribution of strategic effects that can be induced by the temporal regularity of many reaction time tasks (see, e.g., Grosjean, Rosenbaum, & Elsinger, 2001; Lupker, Brown, & Colombo, 1997).

The list of 168 syllables was presented to each participant 6 times, during 6 separate trial blocks. Stimuli in each of the 30 trial blocks (6 blocks × 5 participants) were presented in a different random order. Each trial block took approximately 45 min, and participants completed 3 trial blocks per day. Participants were instructed to take breaks as needed during trial blocks and were given ample rest periods between trial blocks.

Data Preparation

Acoustic data were recorded onto digital audiotape and then transferred to a Sun workstation at a sampling rate of 20 kHz. Segmentation and labeling of the digitized speech tokens were accomplished by a trained phonetician naive to the purposes of the study using the ESPS/Waves+ (Entropic Research Laboratory, Inc., Washington, DC) speech signal processing package. The first four formant center frequencies and their bandwidths were automatically tracked using the default settings. These formant tracks were inspected for accuracy and hand corrected when necessary. Two acoustic boundaries—the onset of acoustic energy and the acoustic onset of the vowel—were hand marked for each speech token, following the standard criteria used to label the Australian National Database of Spoken Language (e.g., Croot, Fletcher, & Harrington, 1992).

The onset of acoustic energy (excluding lip pops and other nonspeech sounds) was denoted by a clear increase in amplitude on the speech waveform following a period of silence. For syllables beginning with the stop consonants and affricates, the onset of acoustic energy was labeled at the burst-release phase of production, following a period in which the vocal tract is occluded completely and there is no acoustic radiation from the lips (see Halle et al., 1957). Delayed naming latency (from which we can make inferences about the EAI) was calculated as the interval between the onset of the signal to respond and the onset of acoustic energy.

The criteria used to determine the acoustic onset of the vowel were dependent on the phonetic characteristics of the preceding phoneme. When vowels followed stop, affricate, or fricative consonants, the acoustic vowel onset was marked at the onset of visible energy on the spectrogram at F2 and above. When vowels followed /m/, /n/, or /l/, the acoustic vowel onset was labeled at the point of marked increase in intensity visible on the spectrogram (these consonants are characterized by a drop in intensity relative to adjacent segments). Finally, when vowels followed /r/, /w/, or /j/, the acoustic vowel onset was marked at the midpoint of the formant transitions between the consonantal and vowel targets. Targets for /w/ and /r/ were defined by a dip in F2 and F3, and the target for /j/ was defined by F2 and F3 maxima. Targets for /i/ were labeled at F2 peaks; targets for /a/ were labeled at F1 peaks; and targets for the neutral vowel /əː/ were labeled at F1 peaks or, if none, at the vowel midpoint (see, e.g., Watson & Harrington, 1999, and Harrington & Cassidy, 1999, for detailed descriptions of Australian English vowel targets).

Data were inspected for the presence of errors and outliers. A small number of tokens (73 tokens, 1.45% of the data) were judged to be mispronunciations or articulatory errors and were removed. We removed a further 54 outlying tokens (1.07% of the data) produced with latencies greater than three standard deviations above a participant's own mean. Delayed naming latencies and onset acoustic durations (calculated as the interval between the onset of acoustic energy and the acoustic onset of the vowel) were averaged across the six repetitions of each syllable for each of the 5 talkers.

Results

We present three stages of data analysis. In the first stage, we sought to establish the extent to which the EAI and OD data provide a reliable estimate of execution-level influences on speech production. In the second stage, we examined the manner in which three articulatory properties of the initial phoneme (manner of articulation, voicing, and place of articulation) influence EAI and OD. In the third stage of analysis, we investigated the influence of two properties of noninitial phonemes (vowel height/fronting and onset cluster complexity) on EAI and OD. Where possible, we report analyses both by participants and by items.

Stage 1: Reliability and Locus of Effects

EAI and OD data for each of the 168 syllables, averaged across the 5 talkers, appear in the Appendix.[1] We assessed the reliability of the EAI and OD data through the use of intraclass correlation analyses (e.g., Shrout & Fleiss, 1979). These analyses revealed a very high degree of reliability for the average of the 5 talkers, both for the EAI data, $r_{(5\ \text{talkers})} = .90$, $p < .01$, and for the OD data, $r_{(5\ \text{talkers})} = .98$, $p < .01$. The reliability of the EAI data is improved further by averaging across the three vowel types, $r_{(5\ \text{talkers})} = .95$, $p < .01$. These figures indicate that variance in the data reported in the Appendix is due largely to true differences in EAI and OD across the 168 syllables and not due to measurement error.

To obtain a coarse estimate of the extent to which our measures reflect the speech motor execution level of processing, we used analysis of variance (ANOVA) to examine the contributions of two basic articulatory parameters—manner of articulation (four levels: stops/affricates, fricatives, nasals, and liquids/semivowels) and voicing (two levels: voiced and unvoiced)—to variance in the EAI and OD data contained in the Appendix. Into the analysis of OD we added the number of onset phonemes (1, 2, or 3), because this is likely to be a major determinant of onset acoustic duration. These factors alone accounted for a remarkable portion of the variance in the EAI ($r^2 = .90$) and OD ($r^2 = .93$) data. We consider these figures to be consistent with the claim that variance in the EAI and OD data reported in the Appendix can be attributed to an execution level of processing.

Stage 2: Initial Phoneme Influences on EAI and OD Data

Table 1 provides a summary of EAI and OD differences across initial phonemes. All values in Table 1 were averaged across the 5 talkers and across the three vowel types. EAI values for each initial phoneme were further averaged across all of the onsets comprising that initial phoneme (e.g., the latency value for /f/ includes data from syllables comprising the onsets /f/, /fr/, and /fl/). Complex onsets (e.g., /fr/, /skl/) are not distributed evenly across the initial phonemes, and one may therefore be concerned about these averaging operations. However, analyses in Stage 3 establish that effects of onset complexity on the EAI are small, statistically marginal, and restricted to stops. The number of phonemes in an onset does, however, have a very strong effect on OD, as revealed by analyses in Stage 1, $F(2, 165) = 245.77$, $p < .01$, $MSE = 313.53$. Thus, OD values in Table 1 reflect only those onsets comprising a single phoneme (e.g., the onset duration value for /s/ includes data from syllables comprising the /s/ onset only).

Table 1 reveals significant consonantal influences on EAI and OD, with average latencies ranging from 223 ms to 302 ms and average durations ranging from 16 ms to 161 ms. In this section, we establish the way in which three articulatory characteristics of initial phonemes—manner of articulation, voicing, and place of articulation—contribute to variance in the EAI and OD data contained in Table 1. In each of these analyses, we averaged across vowel type. Further, for reasons explained above, EAI analyses averaged across onset complexity, whereas OD analyses were restricted to syllables comprising single-phoneme onsets. Finally, affricates were excluded from the following analyses because they comprise aspects of both stop and fricative articulation.

Manner of articulation. We investigated the influence of manner of articulation of the initial phoneme (four levels: stop, fricative, liquid/semivowel, and nasal) on EAI and OD measures through by-subject and by-item ANOVAs. Manner of articulation was treated as a repeated factor in the by-subjects analyses and as an unrepeated factor in the by-items analyses. Mean EAI and OD data, by subjects, are shown in Table 2.

The ANOVAs revealed significant effects of manner of articulation on EAI, $F_1(3, 12) = 32.45$, $p < .01$, $MSE = 101.28$, and

---

[1] As we indicated earlier, the interval between signal onset and acoustic onset (i.e., delayed naming latency) includes not only the EAI but also the signal decoding time, the unit retrieval time, and the unpacking time. Thus, the EAI values in the Appendix are overestimates of the absolute values of the EAI. However, the differences between the EAI values in the rows of the Appendix can be ascribed solely to differences in the EAI, as argued earlier.

Table 1
EAI and OD Summary Data for Each Initial Phoneme Averaged Across Vowel Type

Initial phoneme	EAI (ms)	OD (ms)
/ʃ/ (shut)	223	157
/s/ (sat)	225	161
/f/ (fat)	236	120
/θ/ (thing)	242	129
/z/ (zip)	247	121
/h/ (hat)	248	89
/m/ (man)	255	70
/j/ (yam)	258	91
/n/ (nut)	261	69
/ð/ (that)	263	77
/l/ (lamb)	268	74
/r/ (rat)	268	85
/v/ (vat)	270	73
/w/ (was)	274	79
/dʒ/ (junk)	280	63
/tʃ/ (chip)	285	100
/k/ (key)	287	80
/t/ (tip)	288	75
/p/ (pet)	291	64
/b/ (bat)	301	16
/d/ (dark)	302	23
/g/ (got)	302	27

Note. EAI values for each initial phoneme are averaged across onset complexity. OD values for each initial phoneme comprise single-phoneme onsets only. EAI = execution–acoustic interval; OD = onset acoustic duration.

$F_2(3, 50) = 130.23$, $p < .01$, $MSE = 116.51$, and on OD, $F_1(3, 12) = 31.97$, $p < .01$, $MSE = 129.53$, and $F_2(3, 16) = 7.19$, $p < .01$, $MSE = 767.23$. For the EAI measure, these main effects reflected longer delayed naming latencies for stops than for fricatives, $t_1(4) = 6.23$, $p < .01$, and $t_2(46) = 19.71$, $p < .01$; nasals, $t_1(4) = 4.46$, $p < .01$, and $t_2(20) = 4.60$, $p < .01$; and liquids/semivowels, $t_1(4) = 5.19$, $p < .01$, and $t_2(22) = 4.71$, $p < .01$, and shorter delayed naming latencies for fricatives than for nasals, $t_1(4) = 9.76$, $p < .01$, and $t_2(28) = 3.20$, $p < .01$, and liquids/semivowels, $t_1(4) = 6.21$, $p < .01$, and $t_2(30) = 5.94$, $p < .01$. For the OD measure, these main effects reflected longer onset durations for fricatives than for stops, $t_1(4) = 6.12$, $p < .01$, and $t_2(12) = 4.57$, $p < .01$, and marginally longer onset durations for fricatives than for nasals, $t_1(4) = 6.56$, $p < .01$, and $t_2(8) = 2.05$, $p < .06$, and liquids/semivowels, $t_1(4) = 6.80$, $p < .01$, and $t_2(10) = 1.93$, $p < .07$.

Table 2
Influence of Manner of Articulation of the Initial Phoneme on EAI and OD

	EAI (ms)		OD (ms)	
Manner class of initial phoneme	*M*	*SD*	*M*	*SD*
Stop	295	49	47	5
Liquid/semivowel	267	41	82	17
Nasal	259	40	70	16
Fricative	233	41	116	27

Note. OD values are restricted to syllables comprising single-phoneme onsets (e.g., /sa/). EAI = execution–acoustic interval; OD = onset acoustic duration.

The data indicate that the EAI was shortest for fricatives and longest for stops, with nasals, liquids, and semivowels falling in between. OD was longest for fricatives and shortest for stops, with nasals, liquids, and semivowels falling in between. These observations concerning EAI and OD are entirely consistent with the characterization of stops described in the introduction to this article. Because acoustic energy is not present during the articulatory closure in the production of stops (in contrast to other manner classes of articulation), stops should yield particularly long EAIs and particularly short ODs relative to other manner classes of consonants.

Voicing. We next investigated the influence of voicing on EAI and OD within the manner classes that allow a voicing contrast: stops and fricatives. In the class of fricatives, we considered only those exemplars that allow a voicing contrast (/s z/, /f v/, and /θ ð/). We conducted ANOVAs on the delayed naming latency and onset duration data that treated voicing (two levels: voiced vs. unvoiced) and manner class (two levels: stop vs. fricative) as factors. These factors were treated as repeated in the by-subjects analysis and as unrepeated in the by-items analysis. Mean EAI and OD data, by subjects, are shown in Table 3.

Analyses of the EAI data revealed a main effect of manner class of articulation with fricatives producing shorter delayed naming latencies than stops, $F_1(1, 4) = 26.00$, $p < .01$, $MSE = 490.38$, and $F_2(1, 44) = 338.02$, $p < .01$, $MSE = 52.04$; a main effect of voicing with voiceless exemplars producing shorter delayed naming latencies than voiced exemplars, $F_1(1, 4) = 72.97$, $p < .01$, $MSE = 34.08$, and $F_2(1, 44) = 64.36$, $p < .01$, $MSE = 52.04$; and a Manner Class × Voicing interaction with the effect of voicing reduced for the stops, $F_1(1, 4) = 16.95$, $p < .01$, $MSE = 24.43$, and $F_2(1, 44) = 10.41$, $p < .01$, $MSE = 52.04$. The effect of voicing for the class of stops remained significant, however, $t_1(4) = 4.11$, $p < .05$, and $t_2(18) = 5.73$, $p < .01$.

Analyses of the OD data revealed a main effect of manner class of articulation with fricative onsets being longer in acoustic duration than stop onsets, $F_1(1, 4) = 34.16$, $p < .01$, $MSE = 694.93$, and $F_2(1, 10) = 26.27$, $p < .01$, $MSE = 507.95$, and a main effect of voicing with voiceless onsets being longer in acoustic duration than voiced onsets, $F_1(1, 4) = 145.59$, $p < .01$, $MSE = 89.68$, and $F_2(1, 10) = 13.85$, $p < .01$, $MSE = 507.95$. There was no interaction between these two factors in the OD corpus (both Fs < 1).

Table 3
Influence of Voicing of the Initial Phoneme on EAI and OD for Fricatives and Stops

	EAI (ms)		OD (ms)	
Manner class and voicing	*M*	*SD*	*M*	*SD*
Fricative				
Voiced	260	45	90	27
Unvoiced	228	40	142	30
Stop				
Voiced	301	49	22	6
Unvoiced	288	48	73	13

Note. OD values are restricted to syllables comprising single-phoneme onsets (e.g., /sa/). EAI = execution–acoustic interval; OD = onset acoustic duration.

Place of articulation. We next investigated the influence of place of articulation and its interaction with voicing on EAI and OD, for fricative and stop onsets separately. In the analysis of fricatives, we considered only those exemplars that allow a voicing contrast /s z f v θ ð/. Although by-subjects and by-items analyses were conducted on the EAI data, there were insufficient degrees of freedom to conduct a by-items analysis on the OD data (recall that this analysis examined only single-phoneme onsets). Mean EAI and OD values for the six fricative onsets (/s z f v θ ð/) and the six stop onsets (/p b t d k g/) under investigation can be found in Table 1.

In the analysis of fricatives, we conducted 3 × 2 ANOVAs on the delayed naming latency and onset duration data, treating place of articulation (three levels: alveolar /s z/, interdental /θ ð/, and labiodental /f v/) and voicing (two levels: voiced and unvoiced) as repeated factors in the by-subjects analysis and as unrepeated factors in the by-items analysis. The analysis of EAI revealed a significant effect of place of articulation, $F_1(2, 8) = 5.48$, $p < .05$, $MSE = 168.23$, and $F_2(2, 19) = 25.96$, $p < .01$, $MSE = 12.33$; a significant effect of voicing, $F_1(1, 4) = 55.88$, $p < .01$, $MSE = 88.42$, and $F_2(1, 19) = 128.08$, $p < .01$, $MSE = 12.33$; and no interaction between these two factors, $F_1 < 1$, and $F_2(2, 19) = 3.09$, $p < .10$. Further analysis revealed that the effect of place of articulation reflected shorter delayed naming latencies for the alveolar fricatives /s z/ than for the interdental fricatives /θ ð/, $t_1(4) = 3.25$, $p < .05$, and $t_2(19) = 5.96$, $p < .01$, and the labiodental fricatives /f v/, $t_1(4) = 2.21$, $p < .10$, and $t_2(6) = 6.27$, $p < .01$. The analysis of OD also revealed a significant effect of place of articulation, $F_1(2, 8) = 11.93$, $p < .01$, $MSE = 487.52$; a significant effect of voicing, $F_1(1, 4) = 123.99$, $p < .01$, $MSE = 128.37$; and no interaction between these two factors, $F_1(2, 8) = 1.11$, *ns*. The pattern of onset duration effects was in the reverse direction to the latency effects, with the alveolar fricatives /s z/ being longer in acoustic duration than the interdental fricatives /θ ð/, $t_1(4) = 3.72$, $p < .05$, and the labiodental fricatives /f v/, $t_1(4) = 4.04$, $p < .05$.

In the analysis of stops, we conducted 3 × 2 ANOVAs on the delayed naming latency and onset duration data, treating place of articulation (three levels: bilabial /p b/, alveolar /t d/, and velar /k g/) and voicing (two levels: voiced and unvoiced) as repeated factors in the by-subjects analysis and as unrepeated factors in the by-items analysis. The analysis of EAI revealed only an effect of voicing, $F_1(1, 4) = 16.69$, $p < .05$, $MSE = 74.38$, and $F_2(1, 14) = 26.36$, $p < .01$, $MSE = 31.08$; there was neither an effect of place of articulation (both Fs < 1) nor an interaction between place of articulation and voicing, $F_1(2, 8) = 1.14$, *ns*, and $F_2 < 1$, on stop latencies. The analysis of OD revealed a significant effect of place of articulation, $F_1(2, 8) = 12.65$, $p < .01$, $MSE = 36.07$, and voicing, $F_1(1, 4) = 49.21$, $p < .01$, $MSE = 395.37$, with no interaction between these factors, $F_1(2, 8) = 1.42$, *ns*. Further investigation revealed that the bilabial stops /p b/ had shorter acoustic durations than the alveolar stops /t d/, $t_1(4) = 4.48$, $p < .05$, and the velar stops /k g/, $t_1(4) = 3.99$, $p < .05$, replicating previous findings (e.g., Lisker & Abramson, 1964).

Stage 3: Noninitial Phoneme Influences on EAI and OD Data

Analyses in Stage 2 revealed that articulatory characteristics of initial phonemes have a pronounced effect on EAI and OD in CV syllables. In this section, we consider whether characteristics of noninitial phonemes also have an influence on these periods.

Vowel height and fronting. We first examined the influence of vowel type on EAI and OD. We conducted by-subjects and by-items ANOVAs on the delayed naming latency and onset duration data, treating vowel type (three levels: /ɔ:/, /a/, and /i/) as a repeated factor in both cases. We restricted this analysis to syllables with single-phoneme onsets because (as explained below) we believed that any influence of the vowel on latencies would be most pronounced in these cases.

Analysis of the OD data replicated previous findings (Klatt, 1975; Schwartz, 1969): Consonant acoustic duration was influenced by the characteristics of the subsequent vowel, $F_1(2, 8) = 3.40$, $p = .08$, $MSE = 22.05$, and $F_2(2, 42) = 9.81$, $p < .01$, $MSE = 31.72$. The acoustic durations of consonants preceding the high vowel /i/ were greater (87 ms) than those of consonants preceding the low vowel /a/ (80 ms), $t_1(4) = 2.86$, $p < .05$, and $t_2(21) = 4.39$, $p < .01$. The acoustic durations of consonants preceding the neutral vowel /ɔ:/ fell between those of consonants preceding /a/ and /i/ (85 ms) but did not differ reliably from either.

Overall, an influence of vowel type was not observed on the EAI: Delayed naming latencies for syllables comprising /a/, /ɔ:/, and /i/ vowels were statistically equivalent (266 ms, 269 ms, and 270 ms, respectively), $F_1 < 1$, and $F_2(2, 42) = 1.44$, *ns*. However, it is important to examine the influence of vowel type on the EAI separately for consonants characterized by a complete occlusion of the vocal tract (i.e., stops and affricates). For such consonants, it is possible that the durational influences of vowel type on OD that we observed (see above) affected the closure period itself. Specifically, the closure period may have been reduced in the low-vowel (/a/) context. Because the acoustic correlate of the closure interval for the syllable–initial stops and affricates used in this study was silence, we might expect a compressed EAI (i.e., faster latencies) for syllables comprising low vowels (e.g., /ba/) than high vowels (e.g., /bi/). Our analyses revealed a numerical influence of vowel type on stop and affricate latency in the predicted direction (/a/: 291 ms, /ɔ:/: 297 ms, and /i/: 299 ms), but this influence was not statistically reliable, $F_1 < 1$, and $F_2(2, 14) = 1.51$, *ns*.

Onset cluster complexity. As explained earlier, the number of phonemes in an onset has a strong effect on OD. Because onset cluster complexity is confounded with number of phonemes, we did not investigate its influence on OD.

However, previous research concerning the influence of cluster complexity on the acoustic durations of individual consonants comprising an onset provides some reason to believe that this variable might influence the EAI. In a phenomenon known as compensatory shortening (Lindblom & Rapp, 1973), the acoustic durations of individual consonants comprising an onset are shorter when they are produced in complex contexts (e.g., /sla/) than in singleton contexts (e.g., /sa/; e.g., Umeda, 1977). In the case of clusters containing stops (e.g., /pla/, /spa/), one possibility is that the shortening of consonants in clusters extends to the closure phase—the acoustic correlate of which is silence (at least for the stops produced in this study). For syllables beginning with stop consonants, the effect of closure compression due to onset complexity would be a compression of the EAI.

We thus investigated the influence of onset cluster complexity on the EAI by conducting 2 × 2 ANOVAs on delayed naming latencies, which treated onset complexity (two levels: complex vs. simple) and manner class (two levels: stop vs. fricative) as repeated factors in the by-subjects analysis and as repeated and unrepeated factors, respectively, in the by-items analysis. The analysis was restricted to the stops and fricatives /p b t d k g θ ʃ s f/ because only these initial phonemes can occur in both singleton and complex contexts. Results revealed the expected numerical interaction between onset complexity and manner class: An onset complexity advantage emerged for syllables beginning with stop consonants (simple: 300 ms, complex: 293 ms) but was absent for syllables beginning with fricative consonants (simple: 231 ms, complex: 232 ms). Although this interaction was highly significant in the by-items analysis, it was not reliable by subjects, $F_1(1, 4) = 1.87$, *ns*, and $F_2(1, 8) = 16.27$, $p < .01$, $MSE = 5.22$. The lack of significance in the by-subjects analysis was due to a single talker who did not show the expected pattern.

General Discussion

In this research, we sought to contribute to the development of a temporal characterization of the motor execution stage of speech production by examining consonantal effects on delayed naming latency (which, we argue, must be attributed to the EAI) and on onset acoustic duration (OD). Our aims were twofold: (a) to facilitate greater precision in experiments that require naming latency or duration measures and (b) to contribute to a specification of the speech motor execution system that gives rise to those consonantal effects. To this end, we took hand-labeled delayed naming latency and onset acoustic duration measurements of the production of prepared syllables, when these syllables comprised all possible consonantal onsets of English in three vowel contexts (e.g., /səː/, /sa/, and /si/). Results revealed substantial differences in delayed naming latency (i.e., EAI) and OD across the consonantal onsets of English. Our analyses confirmed that these data are highly reliable and likely reflect variance that arises at a speech motor execution level of processing.

Implications for Naming Experiments

Inferences based on naming latency measurements (and, to an increasing extent, acoustic duration measurements) have played a central role in the development of current theories regarding the cognitive processes underpinning the derivation of phonology from print and conceptual knowledge (e.g., Coltheart et al., 2001; Levelt, Roelofs, & Meyer, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996). Chronometric models involving spoken utterances pose special challenges, however, because of the need to control relevant aspects of the motor response itself, which differs on every trial. Of particular concern has been the possibility that the interval between the onset of motor execution and the onset of acoustic energy (i.e., the EAI)—a component of the naming latency measure—may differ across speech onsets. After all, the onset of acoustic energy is an arbitrary point in the speech production process, favored in early speech chronometry not because of its special significance for cognition but because of practical issues related to its measurement (see Rastle & Davis, 2002). Our findings substantiate this concern. We observed consonantal (and, to a lesser extent, postconsonantal) influences on delayed naming latency, which we attribute to the EAI, spanning a range that makes even the most robust of linguistic effects (e.g., the frequency effect; see Schilling, Rayner, & Chumbley, 1998) pale in comparison. Consonantal and postconsonantal effects of a similar magnitude emerged in the measure of OD.

These findings confirm the need to exercise careful control over the articulatory properties of spoken utterances, aspects of which may be used to inform theories of cognition. They also provide material from which clear recommendations regarding stimulus matching can be drawn. For measures of naming latency, our data confirm that stimulus matching on initial phoneme is necessary. However, for syllables beginning with oral stops, we further recommend matching on the complete syllabic onset. This recommendation is based on the marginal influence of onset cluster complexity (i.e., the complexity advantage) on stop latency. For measures of onset duration, our findings demonstrate that stimulus matching on the complete syllabic onset plus the vowel (e.g., /pliz/ vs. /plid/) is required in all cases. We do not yet know whether or how coda phonemes influence the EAI and OD periods. Accordingly, these recommendations present minimum standards. A more desirable strategy is to use within-target comparisons, in which a particular speech response can be compared with itself, whenever possible (e.g., Kello & Plaut, 2000; Lupker et al., 1997). In the event that the recommended matching standards cannot be achieved for all items in an experiment, the data presented in this article (Table 1 and the Appendix) may be consulted to provide an estimate of the likely consequences of the matching failure and/or to guide the selection of alternative items.

It is important to emphasize that the methodological challenges discussed in this article are quite separate from those introduced by the use of an electronic voice key to measure naming latency or acoustic duration. We have identified articulatory contributions to latency and duration measures, which persist even if these measures are obtained manually via visual inspection of the acoustic waveform or spectrogram. The use of a voice key to detect acoustic boundaries poses an additional set of problems for naming research. It is well known that voice keys are highly inaccurate—in some cases, introducing error approaching 150 ms (Rastle & Davis, 2002) to naming latency measures. However, because the degree of voice key error depends on a phoneme's amplitude, it affects the detection of some phonemes more than others (Sakuma et al., 1997). Thus, a voice key study of naming latency that exercised no control over the phonetic properties of speech responses would be contaminated by variance from both articulatory factors and voice key error—with these sources of contamination often working in opposing directions (e.g., relative to other phonemes, acoustic energy for /s/ is generated quickly but is detected poorly by a voice key; see Rastle & Davis, 2002).

An investigator might be tempted to believe that he or she can solve both of these problems by adopting the stimulus-matching guidelines described above. However, this conclusion would be erroneous. In their study of voice key error, Rastle and Davis (2002) reported that the two types of voice key that they studied failed to detect the initial (voiceless) consonants of a syllable approximately 20% and 60% of the time—not triggering until the acoustic vowel onset. When this occurs, the duration of the initial consonant or consonants becomes a component of the latency measure. Rastle and Davis thus recommended stimulus matching on syllabic onset whenever a voice key is used. However, our present findings demonstrate that the acoustic duration of initial consonants is also influenced by the characteristics of the following vowel (see also, e.g., Klatt, 1975; Schwartz, 1969). This observation suggests that the use of a voice key demands even more stringent phonetic matching across conditions than Rastle and Davis suggested—at a minimum, the syllabic onset plus the vowel.

Implications for Understanding Speech Motor Execution

The data that we have reported are of interest not only with regard to the methodological issues described above but also with regard to understanding speech motor execution. If the consonantal and postconsonantal differences in EAI and OD that we observed can be ascribed to a motor execution level of processing—and we have argued that they can be—then we should be able to account for these differences in terms of specific constraints characterizing that stage of processing. In this section, we therefore consider the observed EAI and OD effects in light of what is known about speech motor execution, seeking to develop testable hypotheses about the biomechanical constraints that give rise to those effects. Because EAI data have so rarely been considered with respect to this issue, we focus in particular on the articulatory properties of initial phonemes that accounted for variance in that corpus—manner of articulation, stop and fricative voicing, and place of fricative articulation—all of which also accounted for variance in the OD corpus.

Manner of Articulation

Syllables beginning with stop consonants had longer delayed naming latencies (i.e., EAIs) and shorter ODs than syllables beginning with other manner classes of consonants. Unlike these other manner classes of consonants in which air is allowed to pass continuously across a constriction (thus generating acoustic energy), one phase of stop production consists of a complete occlusion of the vocal tract. This closure phase (which can last approximately 75–125 ms in voiceless stops) is characterized by a significant rise in intraoral air pressure that is dropped sharply in the next phase of stop production, when the closure is released. At this time, airflow across the oral constriction increases rapidly, and acoustic energy becomes present in the form of a burst (see, e.g., Halle et al., 1957; Löfqvist & Gracco, 1994, 1997, for acoustic and articulatory descriptions of stop-consonant production). The mechanics of stop-consonant production may thus explain why in our delayed naming data acoustic energy for syllables beginning with stop consonants emerged not only later but also over a shorter duration than acoustic energy for syllables containing nonstop initial consonants.

Voicing

Syllables beginning with voiced consonants, be they stops or fricatives, had longer delayed naming latencies (i.e., EAIs) and shorter ODs than syllables beginning with unvoiced consonants. Voicing requires a careful balance of opposing abductory and adductory forces at the glottis (forces acting to open and close the vocal folds), associated with complex patterns of activity in the muscles of the larynx. A possible explanation for the observed effect of voicing on the EAI is that extra time is required (from the onset of motor execution) in the production of voiced initial sounds to achieve this careful balance. This explanation would say nothing, however, about why voiced onsets were of shorter acoustic duration than unvoiced onsets, and we prefer an account in which both latencies and onset durations are systematically related to the differing physiological demands within the vocal tract of voiced and unvoiced initial phonemes. We therefore consider other potential explanations for the effect of voicing, separately for syllables beginning with stops and with fricatives.

Stops. We believe that the inverse relation between EAI and OD in the production of voiced and voiceless stops may be related to properties of the stop closure interval, which precedes the release. Voiceless stops exhibit closure durations that are longer (e.g., Byrd, 1993; Lisker, 1957) and that have a greater mean intraoral air pressure (see Ladefoged & Maddieson, 1996) than voiced stops. These characteristics of the closure may result in a pressure drop on release that not only is larger (e.g., Lubker & Parris, 1970) but also decays over a longer period of time for the voiceless exemplars than for the voiced exemplars (Lisker, 1970; see also Klatt, 1975)—hence, accounting for the longer acoustic duration of the voiceless stops (i.e., their longer voice onset time; Lisker & Abramson, 1964). Despite their typically longer closure duration, however, we observed shorter delayed naming latencies for the voiceless stops than for the voiced stops (see also Fowler, 1979). This apparent contradiction might be reconciled by considering evidence (also considered by Fowler, 1979) that the articulatory approach to closure is often more rapid for the voiceless stops than for the voiced stops (e.g., Smith & McLean-Muse, 1987; see also Ladefoged & Maddieson, 1996). This effect of voicing on the velocity of closing gestures may work to counteract the greater intraoral air pressures involved in the production of voiceless stops (MacNeilage & Ladefoged, 1976).

Fricatives. The production of the fricatives /s z f v θ ð/ requires the generation of turbulence in the airflow downstream from a narrow supraglottal constriction. To generate a turbulent airflow, a talker must achieve a degree of intraoral air pressure (relative to atmospheric pressure) that is sufficient to effect a critical volume flow rate of air across the supraglottal constriction (Stevens, 1971). We believe that the inverse relation between EAI and OD in the production of voiceless and voiced fricatives may be a consequence of the additional aerodynamic demands introduced by the glottal constriction necessary for voicing. When there is no glottal constriction (as in the production of voiceless fricatives), the subglottal pressure is dropped almost entirely across the supraglottal constriction (Stevens, 1971). However, when a glottal constriction is present (as in the production of voiced fricatives), the subglottal pressure is dropped across both constrictions. The consequence of the glottal constriction is thus a reduced pressure drop across the supraglottal constriction—and, hence, a reduction in the volume flow rate of air across this constriction (see Stevens, 1971, Figure 11). In such circumstances, an increase in subglottal pressure may be required to ensure that the volume flow rate across the supraglottal constriction is sufficient to produce turbulence. Such an increase in subglottal pressure would take time to achieve, perhaps accounting for the effect of voicing on the EAI that we observed. Voiced fricatives may also be more difficult to sustain than voiceless fricatives, the reason being that the high intraoral pressures required to generate a turbulent airstream make it difficult to maintain the transglottal pressure differential necessary for voicing (which must be at least 2 cm H_2O; Catford, 1977).

Place of Fricative Articulation

Syllables beginning with the alveolar fricatives /s z/ had longer delayed naming latencies (i.e., EAIs) and shorter ODs than syllables beginning with either the interdental fricatives /θ ð/ or the labiodental fricatives /f v/. We believe that this influence of place of articulation may be traced to two fundamental aerodynamic properties of generating and maintaining a turbulent airflow in fricative production. First, recent magnetic resonance imaging data have shown that the alveolar fricatives /s z/ have a much smaller supraglottal constriction area than the labiodental and interdental fricatives (which have very similar constriction areas; Narayanan et al., 1995). The critical volume flow rate of air required to generate turbulence depends on the supraglottal constriction area, such that an increased supraglottal constriction area requires an increased volume flow rate (and, hence, an increased pressure differential across the supraglottal constriction; Stevens, 1971). The production of the labiodental and interdental fricatives may thus require an increase in subglottal pressure relative to the alveolar fricatives. As in the case of voicing, this increase in pressure may take time to achieve and may be difficult to sustain. Second, in the production of the alveolar fricatives /s z/, a jet of air is forced through a narrow supraglottal constriction before striking a secondary obstacle in the flow—namely, the front teeth. The teeth act as a spoiler in the airstream and indeed amplify significantly the turbulence created at the supraglottal constriction (Shadle, 1985). This second difference between the alveolar fricatives and the other classes of fricatives may also contribute to the reduced aerodynamic demands of generating and maintaining a turbulent flow at the supraglottal constriction for /s z/ and may therefore play a role in the latency and duration effects that we observed.

Conclusion

This study quantified consonantal and postconsonantal effects on two intervals within the execution phase of speech production—the execution–acoustic interval (i.e., the EAI) and the onset acoustic duration interval (i.e., the OD). Through principled task analysis (Monsell, 1986; Sternberg et al., 1978), we have attempted to eliminate the influence of preexecution processes on our measures and have argued that our findings may contribute to a fuller understanding of speech motor execution. The connection between these execution-level speech processes and higher level cognitive speech-planning operations has not escaped our attention, however. We believe that speakers plan their articulatory movements prior to the onset of articulation (Rastle, Harrington, Coltheart, & Palethorpe, 2000) and do so on the basis of knowledge about the biomechanical characteristics of the speech apparatus. For example, timing specifications in speech motor plans for fricatives may denote a reduced duration in circumstances in which aerodynamic demands are greater (e.g., when they are voiced). Alternatively, the plan may specify that an action is to be taken at a point where subglottal or intraoral air pressure falls to a specified level, in order to conserve a sufficient volume of air to meet the requirements for upcoming sounds. Either way, we argue that a full understanding of cognitive planning for speech cannot be achieved without also understanding the properties of speech motor execution. Further interdisciplinary research in cognitive psychology, speech physiology, and phonetics is essential for understanding both of these domains of speech production.

References

Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1289–1302.

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.

Baum, S. R. (1996). Fricative production in aphasia: Effects of speaking rate. *Brain and Language, 52,* 328–341.

Baum, S. R., & Blumstein, S. E. (1987). Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *Journal of the Acoustical Society of America, 82,* 1073–1081.

Baum, S. R., & Ryan, L. (1993). Rate of speech effects in aphasia: Voice onset time. *Brain and Language, 44,* 431–445.

Behrens, S., & Blumstein, S. E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *Journal of the Acoustical Society of America, 84,* 861–867.

Bernard, J. R. L. (1967). Length and the identification of Australian English vowels. *Journal of the Australasian Universities Language and Literature Association, 27,* 37–58.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics, 83,* 97–116.

Catford, J. C. (1977). *Fundamental problems in phonetics.* Edinburgh, Scotland: Edinburgh University Press.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1197–1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). The DRC model: A model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

Croot, K., Fletcher, J., & Harrington, J. (1992). Levels of segmentation and labelling in the Australian National Database of Spoken Language. *Proceedings of the Fourth Australian International Conference on Speech Science and Technology,* 86–90.

Fowler, C. A. (1979). "Perceptual centers" in speech production and perception. *Perception & Psychophysics, 25,* 375–388.

Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say "blah!": Attentional dynamics of delayed naming. *Journal of Memory and Language, 37,* 190–216.

Gracco, V. L., & Löfqvist, A. (1994). Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience, 14,* 6585–6597.

Grosjean, M., Rosenbaum, D. A., & Elsinger, C. (2001). Timing and reaction time. *Journal of Experimental Psychology: General, 130,* 256–272.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29,* 107–116.

Harrington, J., & Cassidy, S. (1999). *Techniques in speech acoustics.* Dordrecht, the Netherlands: Kluwer.

Kawamoto, A. H., & Kello, C. T. (1999). Effect of onset cluster complexity in speeded naming: A test of rule-based approaches. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 361–375.

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.

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 26, 719–750.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General, 129,* 340–360.

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

Klatt, D. H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research, 17,* 51–63.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word–initial consonant clusters. *Journal of Speech and Hearing Research, 18,* 389–406.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, England: Blackwell.

Lehiste, I., & Peterson, G. (1961). Transitions, glides, and diphthongs. *Journal of the Acoustical Society of America, 33,* 268–277.

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

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–75.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50,* 239–269.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. In *Papers in linguistics from the University of Stockholm* (Vol. 21). Stockholm: University of Stockholm.

Lisker, L. (1957). Closure duration and the inter-vocalic voiced–voiceless distinction in English. *Language, 33,* 42–49.

Lisker, L. (1970). Supraglottal air pressure in the production of English stops. *Language & Speech, 13,* 215–231.

Lisker, L., & Abramson, A. (1964). A cross language study of voicing in initial stops. *Word, 20,* 384–422.

Löfqvist, A., & Gracco, V. L. (1994). Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context. *Phonetica, 51,* 52–67.

Löfqvist, A., & Gracco, V. L. (1997). Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language and Hearing Research, 40,* 877–893.

Lubker, J. F., & Parris, P. (1970). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electro-myographic activity during the production of stop consonant cognates /p/ and /b/. *Journal of the Acoustical Society of America, 47,* 625–633.

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.

MacNeilage, P., & Ladefoged, P. (1976). The production of speech and language. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7, pp. 76–120). New York: Academic Press.

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.

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

Monsell, S. (1986). Programming of complex sequences: Evidence from the timing of rapid speech and other productions. In H. Heuer & C. Fromm (Eds.), *Generation and modulation of action patterns* (pp. 72–86). Heidelberg, Germany: Springer-Verlag.

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.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America, 98,* 1325–1347.

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

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55*(A), 1339–1362.

Rastle, K., Harrington, J., Coltheart, M., & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1178–1191.

Sakuma, N., Fushimi, T., & Tatsumi, I. (1997). Measurement of naming latency of Kana characters and words based on speech analysis: Manner of articulation of a word–initial phoneme considerably affects naming latency. *Japanese Journal of Neuropsychology, 13,* 126–136.

Savage, G. R., Bradley, D. C., & Forster, K. I. (1989). Word frequency and the pronunciation task: The contribution of articulatory fluency. *Language and Cognitive Processes, 5,* 203–236.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26,* 1270–1281.

Schwartz, M. F. (1969). Influence of vowel environment upon the duration of /s/ and /S/. *Journal of the Acoustical Society of America, 46,* 480.

Schwartz, M. F. (1970). Duration of /s/ in /s/-plosive blends. *Journal of the Acoustical Society of America, 47,* 1143–1144.

Shadle, C. H. (1985). *The acoustics of fricative consonants* (Tech. Rep. No. 506). Cambridge: Massachusetts Institute of Technology.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–427.

Smith, B. L., & McLean-Muse, A. (1987). Kinematic characteristics of postvocalic labial stop consonants produced by children and adults. *Phonetica, 44,* 227–237.

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

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45,* 175–197.

Sternberg, S., Monsell, S., Knoll, R., & Wright, C. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.

Stevens, K. (1971). Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America, 50,* 1180–1192.

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

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America, 61,* 846–858.

Watson, C., & Harrington, J. M. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America, 106,* 458–468.

Appendix

Execution–Acoustic Interval (EAI) and Onset Acoustic Duration (OD) Data (in Milliseconds), for Each Vowel Separately and Averaged Across Vowels

	EAI				OD			
Onset	əː	a	i	*M*	əː	a	i	*M*
b	298	295	316	303	17	14	16	16
bl	295	296	285	292	67	69	71	69
br	313	295	312	307	85	83	83	84
d	314	306	303	307	22	21	26	23
dr	304	299	305	303	101	99	114	104
dw	298	293	298	296	108	103	104	105
g	322	301	304	309	25	25	32	27
gl	308	288	304	300	70	77	80	76
gr	305	285	293	294	110	98	102	103
gw	311	304	295	303	83	88	92	88
p	300	284	305	296	61	66	65	64
pl	279	278	298	285	97	95	99	97
pr	289	301	285	292	117	107	118	114
t	281	294	304	293	71	75	78	75
tr	288	271	293	284	126	129	135	130
tw	289	288	282	286	120	125	127	124
k	280	298	294	291	75	81	84	80
kl	297	282	275	285	97	104	105	102
kr	291	280	282	284	130	125	132	129
kw	289	293	280	287	111	113	119	115
dʒ	285	271	285	280	59	57	74	63
tʃ	295	279	281	285	102	95	104	100
m	253	260	253	255	74	64	71	70
n	255	266	263	261	73	62	73	69
l	283	254	267	268	77	68	76	74
r	273	264	267	268	90	79	87	85
j	264	251	260	258	88	92	94	91
w	274	275	274	274	84	68	84	79
z	253	240	248	247	125	108	131	121
v	254	268	288	270	85	68	65	73
ð	263	262	263	263	82	70	80	77
s	221	239	223	228	162	153	168	161
sk	217	217	228	221	198	185	207	197
skl	224	222	222	223	247	239	245	243
skr	233	233	216	227	267	255	268	263
skw	219	222	232	224	244	233	248	242
sf	224	226	219	223	206	195	209	204
sl	225	233	236	232	202	187	209	199
sm	234	220	221	225	189	189	194	191
sn	223	234	209	222	192	184	190	189
sp	226	228	218	224	206	187	194	195
spl	233	219	212	221	233	220	238	230
spr	223	216	227	222	263	252	275	263
st	223	220	246	230	194	176	199	190
str	207	240	221	223	244	235	262	247
stw	226	238	224	230	248	242	249	246
sw	222	247	219	229	223	206	212	214
f	222	237	236	232	125	109	127	120
fl	231	239	238	236	148	142	155	148
fr	235	253	236	241	179	163	176	172
θ	249	238	242	243	120	132	134	129
θr	236	247	238	241	192	190	197	193
θw	243	242	239	241	206	189	201	199
ʃ	228	218	217	221	159	150	164	157
ʃr	232	212	232	225	229	226	236	230
h	248	247	248	248	87	97	85	89

Received March 8, 2004
Revision received April 1, 2005
Accepted April 11, 2005 ■