Journal of Experimental Psychology:
Human Perception and Performance
2002, Vol. 28, No. 2, 307–314

0096-1523/02/$5.00 DOI: 10.1037//0096-1523.28.2.307

On the Complexities of Measuring Naming

Kathleen Rastle
Macquarie University and University of Cambridge

Matthew H. Davis
Medical Research Council Cognition and Brain Sciences Unit

The aims of this study were to investigate the adequacy of electronic voice keys for the purpose of measuring naming latency and to test the assumption that voice key error can be controlled by matching conditions on initial phoneme. Three types of naming latency measurements (hand-coding and 2 types of voice keys) were used to investigate effects of onset complexity (e.g., *sat* vs. *spat*) on reading aloud (J. R. Frederiksen & J. F. Kroll, 1976; A. H. Kawamoto & C. T. Kello, 1999). The 3 measurement techniques produced the 3 logically possible results: a significant complexity advantage, a significant complexity disadvantage, and a null effect. Analyses of the performance of each voice key are carried out, and implications for studies of naming latency are discussed.

For over a century, experimental psychologists have collected reaction time data, from which they have drawn inferences about the architectural and processing characteristics of cognitive systems (see Meyer, Osman, Irwin, & Yantis, 1988, for a review). With respect to the language processing and speech production systems, these reaction time data have often taken the form of a naming latency measurement—a measurement of the time between the presentation of a target stimulus (be it a written word, picture, spoken word, or sentence) and the onset of a spoken response, at which *onset* is defined acoustically. This measurement is assumed to reflect properties of cognitive processes completed prior to the onset of articulation (e.g., lexical access).[1]

The measurement of the naming latency response has most often been achieved in experimental psychology with the use of an electronic voice key—a hardware device that captures the onset of sound automatically. Indeed, we surveyed the articles published in *Journal of Experimental Psychology: Human Perception and Performance* during the years of 1995–1999 and found that of those that reported collecting naming latencies (37 articles), 95% used a voice key (see also Kessler, Treiman, & Mullenix, in press, who found a similar result in a survey of the articles published in *Journal of Memory and Language* between 1997 and 2000).

Despite the common use of voice keys, however, the issues involved in automatically detecting the acoustic onset of sound are numerous. Indeed, a number of researchers have suggested that voice keys may not reliably detect acoustic onset and have published studies in which voice key error (the time between actual acoustic onset and the point at which the voice key triggers) is examined for different kinds of initial phonemes (see Pechmann, Reetz, & Zerbst, 1989, and Sakuma, Fushimi, & Tatsumi, 1997, both of which found voice key error of over 100 ms for some types of phonemes). Given these reports that voice keys may not reliably detect acoustic onset, especially for particular phoneme classes, experimental psychologists have implemented procedures to ensure that whatever error is associated with the measurement does not vary systematically across the experimental manipulation. The primary means of doing this has been to match experimental conditions on the initial phonemes used. Our aims in this article are twofold. First, we wish to test the assumption that matching experimental conditions on initial phoneme controls voice key error. Second, we wish to provide an examination of voice key error for a common English onset phoneme, /s/, lending some emphasis to the relationship between internal circuitry of the voice key and associated error.

A Brief History of the Voice Key

Since the late 1800s, a variety of devices have been used to capture a physical event representative of the onset of speech, to derive a chronometric measure of response latency. Early devices designed for this purpose, such as the lip key (which detected lip movement) and the breath key (which detected breath expulsion), fell into disuse by the early 1900s, partly because of practical difficulties involved in their use (e.g., physical contact with the lip key), but largely because not all phonemes could be detected equally effectively by these devices (e.g., there is little lip move-

Kathleen Rastle, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, New South Wales, Australia, and Centre for Speech and Language, Department of Experimental Psychology, University of Cambridge, Cambridge, United Kingdom; Matthew H. Davis, Medical Research Council Cognition and Brain Sciences Unit, Cambridge, United Kingdom.

We are grateful to John Baldwin, William Connolly, Ingrid Johnsrude, and Mark Wolfworth for helpful discussion, to Kyra Karmiloff and Mike Cooper for research assistance, and to Brian Cox for assistance with preparing Figure 1.

Correspondence concerning this article should be addressed to Kathleen Rastle, who is now at Department of Psychology, Royal Holloway, University of London, Egham, Surrey TW20 OEX, United Kingdom. E-mail: kathy@maccs.mq.edu.au

[1] The notion that the initiation of articulation requires the generation of a complete phonological code has been the subject of recent controversy (for a discussion of this issue, see e.g., Kawamoto, Kello, Jones, & Bame, 1998; Kello, Plaut, & MacWhinney, 2000; Rastle, Harrington, Coltheart, & Palethorpe, 2000). Thus, although naming latency measurements are assumed to reflect the completion of a number of cognitive events, exactly which events these are has yet to be determined definitively.

ment or expulsion of breath in the phoneme /m/; see Wells & Rooney, 1922). Thus, early in the 20th century, the detection of acoustic energy had become the favored and most common way to measure the naming response, using a mechanical voice-operated switch. These early voice keys (e.g., Dunlap, 1913, 1921; Wells & Rooney, 1922), consisted of a thin metal plate that under no-sound conditions made contact with a nonmagnetic lever or wire. When a participant spoke close to the plate, vibrations produced by the acoustic energy of speech would move it away from the contact lever or wire, breaking an electrical circuit and stopping a timer.

One technological advance that occurred in the development of this class of voice key was the addition of a microphone (Boder, 1933, 1940; Vaughn & Strobel, 1940). Instead of directly breaking a connection, sound vibration would be converted into electrical energy, which, when it exceeded some threshold voltage, would operate a relay. Developments in electronic components saw a move away from the use of electromagnetic relays (e.g., Dunlap, 1913), toward thermionic and gas-filled relays (Fletcher & Bosch, 1938; Guttman, 1957; Kahn, 1935), and then toward the solid-state devices and integrated circuits commonly used in modern voice keys. This additional circuitry between the microphone and the threshold device or switch improved reliability in the detection of acoustic onset. Moreover, it allowed greater sensitivity to energy within particular frequency ranges, or allowed the voice key to respond to the duration as well as to the intensity of sound. Despite these advances, however, all of these systems function in approximately the same fashion; once an acoustic signal exceeds an amplitude threshold, a switch is triggered, stopping the timing clock.

This historical perspective is important insofar as it suggests that the onset of speech can be measured as one of any number of events—the earliest increase in air pressure in the vocal tract, the first expulsion of breath, or the initial movement of the tongue, lips, or jaw. It was, throughout history, the ease and reliability of measuring the start of sound pressure vibration that defined this particular point in the speech signal (the acoustic onset) as one of chronometric importance. Two problems with basing chronometric models on such measurements are immediately apparent, however.

First, there is some evidence to suggest that the generation of acoustic energy is affected by manner of articulation, insofar as different manner classes of phonemes may be realized acoustically at different points in the production process (Fowler, 1979). For example, the production of the phoneme /k/ in *cat* is characterized initially by a complete occlusion of the vocal tract, during which time airflow is prevented and there is no acoustic radiation from the lips. Acoustic energy is emitted in the production of /k/ and other stop consonants only subsequent to this closure period, in the form of a burst created by the disocclusion of the vocal tract (see Halle, Hughes, & Radley, 1957). In contrast, whereas the production of the phoneme /s/ in *sat* requires an oral constriction, the vocal tract is, at no time, occluded completely; air is allowed to pass through the vocal tract throughout the production of /s/, enabling the emission of acoustic energy. Hence, even if the onset of articulation were equivalent for these two words, acoustic energy may be produced earlier for *sat* than for *cat*. Second, even when acoustic energy *is* present, voice keys may not detect it with equal effectiveness for all phonemes. For example, phonemes accompanied by vocal fold vibration, such as the /v/ in *vat,* will be of higher amplitude (and thus more easily detected by a voice key) than phonemes which are not accompanied by vocal fold vibration, such as the /f/ in *fat*. For these two reasons, comparisons of naming latencies for words with different initial phonemes are problematic.

Recent research using the naming task has therefore required that stimuli be matched across conditions on initial phoneme or phonetic class.[2] By matching on initial phoneme, it is assumed by researchers that they can control for (a) the extent to which acoustic energy is *produced* through the articulatory period, and (b) the extent to which voice keys can *detect* any acoustic energy present. As stated earlier, the major focus of this investigation is on the second of these assumptions—that by matching on initial phoneme, researchers assume that any error associated with a voice key will be constant across conditions. We argue here that this assumption may be false, and turn to a recent study, on which our argument is based.

Onset Complexity Effect on Naming Latency

Kawamoto and Kello (1999) have reported that words with complex onsets (e.g., *spat*) are read aloud more quickly than are words with simple onsets (e.g., *sat*). This result is curious, because Frederiksen and Kroll (1976) reported exactly the opposite result when stimuli were also matched for initial phoneme. Kawamoto and Kello suggested that the discrepant results might have been due to different ways of measuring onset latency: Whereas they used a software algorithm to detect the onset of acoustic energy, Frederiksen and Kroll used a voice key.

Kawamoto and Kello (1999) proposed that where they had measured the acoustic onset of each response (using their software algorithm), Frederiksen and Kroll’s (1976) voice key might have responded to a different acoustic event—namely, the onset of voicing. Because some complex items have silence in the second phoneme (e.g., *spat*), the onset of voicing occurs, on average, later for complex words than for simple words. Therefore, if the onset of voicing is measured, a complexity disadvantage may become apparent. This is a situation in which the interval between true acoustic onset and the point detected by a voice key may not be equivalent across conditions, despite matching on initial phoneme.

If different means of measuring naming latency can produce opposite effects of an experimental variable, it would have serious implications for the interpretation of a range of psychological data. In such circumstances, differences in naming latency might not reflect properties of the cognitive processes of interest but would instead reflect discrepancies introduced by response measurement. We therefore set out to investigate how techniques of measuring naming latency can influence response time data, using Kawamoto and Kello’s (1999) onset complexity manipulation as an illustration.

Naming latencies from an experiment in which participants read aloud words with simple and complex onsets were measured in

[2] Eriksen, Pollack, and Montague (1970) advocated the use of the delayed naming paradigm to ensure that voice key error was controlled across the experimental manipulation. However, because response times in delayed naming have been shown to be influenced by nonarticulatory factors (e.g., word frequency; Goldinger, Azuma, Abramson, & Jain, 1997), the use of delayed naming to control for voice key error has largely been abandoned.

three ways: from visual inspection of the speech waveform and with two types of voice key. By recording participants' responses to computer disk, comparisons between different methods of measuring response latency could be made with exactly the same tokens. We examined only words beginning with /s/, primarily because Kawamoto and Kello (1999, Experiment 2) reported a clear complexity advantage using such items. Additionally, because the majority of English monosyllables with complex onsets begin with a voiceless fricative, investigating the characteristics of /s/ in relation to automatic detection is of particular interest. With these data, we investigated two questions: (a) Does the nature of the onset complexity effect in speeded reading aloud vary as a function of measurement technique (comparing hand-coding and two types of voice key), and (b) If so, what interactions between the speech signal and voice key allow this to happen?

Method

Participants

Twenty-four participants between the ages of 18 and 30 years from the University of Cambridge were tested. Participants had normal or corrected-to-normal vision, were native speakers of British English, and were paid £5 (about $7) for their participation.

Materials

Test stimuli were 40 monosyllabic words beginning with the phoneme /s/, half of which contained two-phoneme onsets in which the second phoneme was /p/ or /t/ (e.g., *spat*) and half of which contained one-phoneme onsets (e.g., *sat*). These stimuli were taken from Kawamoto and Kello (1999, Experiment 2); see Kawamoto and Kello for further stimulus description. Forty-four monosyllabic 4–5 letter filler items were included; filler items contained simple or complex onsets and were not /s/ initial. Thus, the complete stimulus set contained a range of initial phonemes, although only the syllables beginning /s/ were of interest in this experiment.

Apparatus

Stimulus presentation and data recording were controlled by the DMDX display system (see http://psy1.psych.arizona.edu/~kforster/dmdx/index.htm) running on a Pentium III PC. Naming responses were recorded directly to the hard drive of the PC at a sampling rate of 22.05 kHz, using a 16-bit SoundBlaster Live! sound card and a Beyerdynamic DT290 microphone (which was attached to a headset intended to keep the microphone at a constant distance from the mouth). Recording began on the presentation of each target and continued for 2 s.

The software voice key contained within DMDX was used to measure naming latency during the experiment (which we refer to as the *simple threshold* voice key). This software voice key monitors the digital output from the sound card (a stream of 16-bit values representing the acoustic signal from the microphone) and records the first time at which an amplitude value (either positive or negative) exceeds a preset threshold. Although this voice key operates within a software package, it is equivalent to a standard, electronic voice key in which a simple amplitude threshold is used to determine the onset of acoustic energy.

A second electronic voice key was also used to measure naming latency from the speech files recorded by DMDX. This voice key (which we refer to as the *integrator* voice key) was designed and constructed by technicians in the electronics laboratory of the Department of Experimental Psychology at the University of Cambridge. Unlike the simple threshold voice key, which is triggered only by a signal that exceeds an amplitude threshold, the integrator voice key is sensitive both to the amplitude of signals and to their duration. In this way, the integrator voice key is triggered not only by signals that exceed a minimum intensity, but also by signals of a lower intensity that continue over time. Because nonspeech signals (e.g., lip pops and clicks) tend to be of short duration, this voice key can be adjusted so that it is more sensitive to low amplitude signals without an increase in the number of triggers to nonspeech sounds.[3] By connecting the output of the computer's sound card to the input of this voice key, we were able to run mock experiments using the recordings of simple and complex words produced earlier by participants.

Procedure

Participants were tested individually in a quiet room. They were fitted with the voice key headset and asked to read a list of 10 words aloud (consisting of a range of initial phoneme types) as the experimenter adjusted the sensitivity of the simple threshold key, so that it did not respond to ambient noise or nonspeech sounds yet did respond to the perceived onset of acoustic energy for the stimulus items. In operational terms, the experimenter attempted to synchronize the perceptual experiences of the heard acoustic onset and a visual signal denoting that the voice key had triggered. The sensitivity of the simple threshold voice key was not readjusted once the experiment had begun.

During the main experiment, each trial proceeded as follows: A fixation cross (+) appeared in the center of the screen for 500 ms and was replaced by a target word presented in lowercase and 28-point font; the target word stayed on the screen until the participant made a response; after a 2-s blank screen, the fixation cross appeared again, signaling the beginning of the next trial. Participants were instructed to read aloud each word as quickly and as accurately as possible, speaking clearly into the microphone. Ten practice trials including words with various initial phonemes preceded the main experiment. Target and filler items were presented in a different random order for each participant.

Subsequent to this testing session, a mock experiment was run using the speech recordings and the integrator key described above. The experimenter adjusted the sensitivity of the integrator voice key for each participant as described above, using the practice items recorded by each participant, once more aiming to synchronize the perceptual experience of the acoustic onset with the signal (in this case a light) denoting voice key detection.

Results

Data from 2 participants were excluded because of equipment failure. Mispronounced tokens were also excluded from the analyses (1.9% of the data). Because of the low number of errors, we did not analyze the error data for effects of onset complexity. The onset of acoustic energy for the target items—denoted by a clear increase in amplitude on the speech waveform following a period of silence (and excluding lip pops or other nonspeech sounds)—was hand-marked from the speech recordings.[4] Voice key data were compared with these hand markings of acoustic onset; voice

[3] In more technical terms, the speech signal is half-wave rectified, amplified, and passed through a leaky integrator circuit (time constant = 0.56 s); the voice key is triggered when the output of this integrator exceeds an amplitude threshold.

[4] Measurements were conducted by Kathleen Rastle. Because the correctness of the onset latency measurements was critical to the aim of the study, Rastle's measurements were verified by a second, naive rater. The correlation between the two sets of markings was .97, and both sets of markings produced a significant 9-ms complexity advantage.

key triggers that preceded the hand-marked acoustic onset (false alarms) were discarded, along with those trials in which the voice key failed to trigger at all (misses). The simple threshold key produced three false alarms (0.35% of the data) and 15 misses (1.73% of the data); the integrator key produced 14 false alarms (1.62% of the data) and no misses.

Three measurements of naming latency for the remaining items are shown in Table 1. It is immediately apparent that both voice keys were triggered some time after the onset of acoustic energy in this experiment. This is perhaps unsurprising given that voiceless fricative onsets are notoriously quiet; indeed, for similar types of phonemes, both Pechmann et al. (1989) and Sakuma et al. (1997) found comparable delays between acoustic onset and voice key detection. However, it has previously been assumed that because the initial segment is matched across each complexity condition, the event detected will be an approximately equal duration after the onset of acoustic energy in both conditions. As shown, this assumption does not appear to be correct for either of the voice keys that were tested in the current study; if it were, both of the voice keys would have revealed a complexity advantage of approximately 9 ms, even though the average naming latencies are greater than those obtained by hand-coding.

These measurements were analyzed statistically by participants and by items, with onset complexity and measurement type as independent factors. In the by-participants analysis, both factors were treated as repeated measures; in the by-items analysis, measurement type was a repeated factor, whereas onset complexity was an unrepeated factor. The interaction between measurement type and complexity was significant, $F_1(2, 42) = 8.46$, $p < .01$, $MSE = 127.11$, and $F_2(2, 76) = 17.46$, $p < .01$, $MSE = 53.60$, as the pattern of complexity effect differed across measurement type. The effect of measurement type also reached significance, $F_1(2, 42) = 79.70$, $p < .01$, $MSE = 2{,}698.20$, and $F_2(2, 76) = 3{,}644.30$, $p < .01$, $MSE = 53.60$, because the three methods of measurement produced different average naming latencies. The main effect of onset complexity did not approach significance, $F_1(1, 21) = 0.004$, and $F_2(1, 38) = 0.06$.

We conducted planned comparisons on the effect of onset complexity in each of the three sets of latency data to investigate the Measurement Type × Onset Complexity interaction. From this single set of data, all three logically possible results emerged. Hand-coded data revealed a significant onset complexity advantage similar to that reported by Kawamoto and Kello (1999), $t_1(21) = 2.89$, $p < .01$, and $t_2(38) = 2.02$, $p = .05$. Data from the simple threshold key showed the reverse effect—a complexity disadvantage similar to that reported by Frederiksen and Kroll (1976), $t_1(21) = -2.17$, $p < .05$, and $t_2(38) = -2.28$, $p < .05$. The onset complexity effect did not approach significance in either direction for the integrator key, $t_1(21) = 0.47$, and $t_2(38) = 0.08$. Thus, it appears that the direction of the complexity effect on naming latency depends on the method used to measure naming latencies.

Table 1
Naming Latencies by Participants as a Function of Onset Complexity and Method of Measurement

Onset complexity	Hand-marking	Simple threshold key	Integrator key
Simple	371	500	449
Complex	362	511	447

To examine why these different results emerged, we marked speech recordings for two further acoustic events: the onset of voicing and the offset of frication. The onset of voicing was determined using the pitch detection algorithm (Talkin, 1995) supplied in the ESPS/Waves+ software package and was defined as being the earliest point at which two successive pitch periods contained voicing.[5] The average onset of voicing was 523 ms from target presentation for simple words and 554 ms for complex words. Thus, the onset of voicing was significantly later for complex words than for simple words, $t_1(21) = 8.99$, $p < .01$, and $t_2(38) = 6.11$, $p < .01$. The duration of the onset (measured from acoustic onset to onset of voicing) was unsurprisingly longer for complex words than for simple words (195 ms vs. 154 ms, respectively), $t_1(21) = 11.27$, $p < .01$, and $t_2(38) = 9.51$, $p < .01$. The offset of frication (the end of the initial /s/ segment) was hand-marked. Initial /s/ segments were shorter when they occurred in a complex context than when they occurred in a simple one (126 ms vs. 154 ms from the onset of acoustic energy, respectively), $t_1(21) = 7.88$, $p < .01$, and $t_2(38) = 7.61$, $p < .01$.

We calculated how far from the onset of acoustic energy each voice key was triggered, for each item, and expressed these values as a proportion of overall onset duration. For example, the simple threshold key recorded a latency of 438 ms for the item *spent* for Participant 1. The hand-coded onset for this item was 341 ms, and the onset of voicing occurred at 549 ms; thus, the duration of the onset for this item was 208 ms. By dividing the duration of the onset into the difference between the voice key and hand-coded measurements (97 ms), one can see that the voice key triggered approximately halfway through the onset of the item (.47). A frequency distribution of these values for the two voice keys is presented in Figure 1, separately for simple and complex words.

Although all items in this experiment contained the same initial phoneme, the point of detection for each voice key ranged from near the onset of acoustic energy into and beyond the onset of voicing. Moreover, the distribution of detection points varied as a function of which voice key was used. Whereas the majority of detections by the simple threshold key were on or after the onset of voicing, the majority of detections by the integrator key occurred during the onset segments of the syllable. This contrast can be explained by referring to the internal circuitry of each voice key. The simple threshold key detects only events in the speech signal that exceed some minimum amplitude, and therefore, in most cases, fails to detect the low-intensity frication characteristic of /s/. The integrator voice key, however, sums the amplitude of acoustic signals over time; it therefore becomes sensitive to the low-intensity fricated portion of the speech signal because it continues over an extended period of time.

The complexity effects revealed by each of the voice keys can be understood as resulting from a balance between detections in onset and voicing segments. The simple threshold key thus showed

[5] A naive rater subsequently hand-marked half the speech recordings for the onset of voicing to determine the validity of the ESPS/Waves+ algorithm. The correlation between these markings and the markings generated by ESPS/Waves+ was .99.

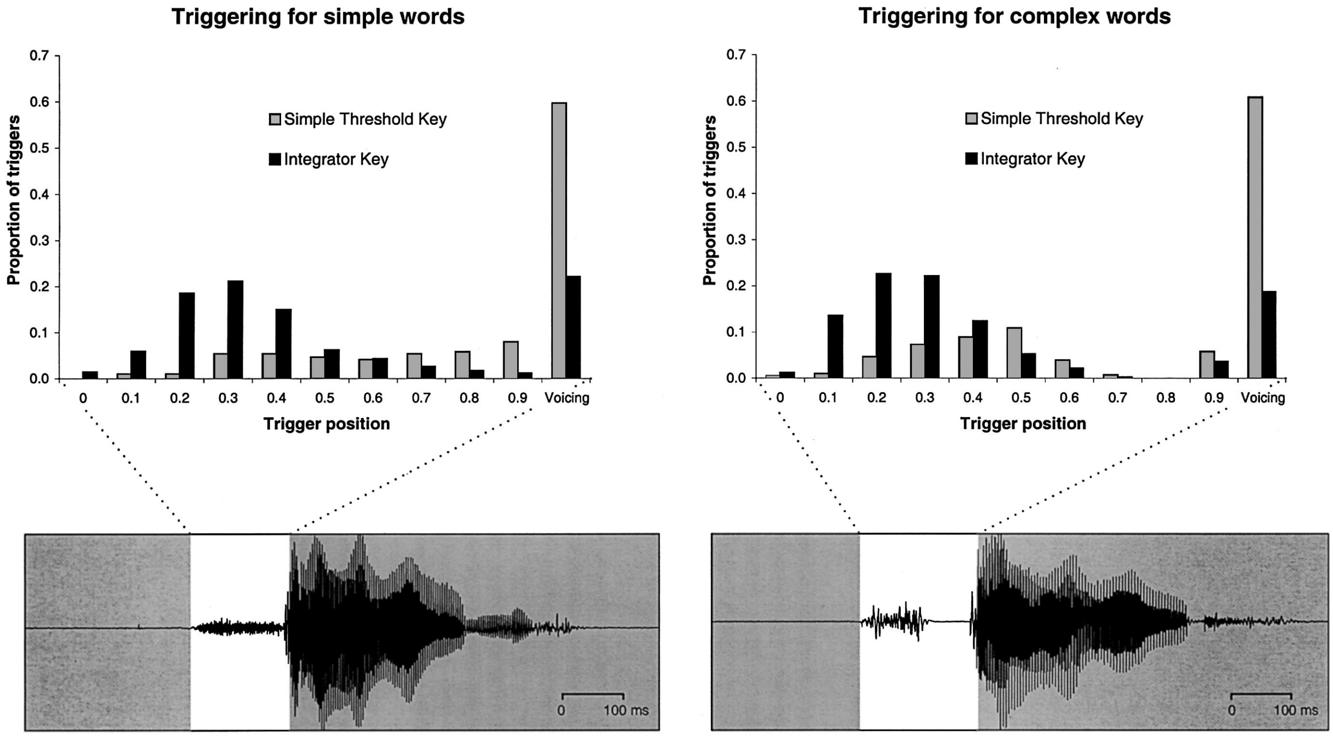

Figure 1. Trigger positions for simple and complex words measured by two different voice keys. Trigger positions normalized with respect to the onset of acoustic energy (0) and the onset of voicing (1). The label *Voicing* refers to all triggers that occur after the onset of voicing.

a small but significant onset complexity disadvantage, detecting approximately 60% of the tokens at or after the onset of voicing, a point that was 31 ms later for complex words than for simple words. For the integrator key, however, the majority of tokens were detected within the onset segment. Although this profile might be expected to produce a complexity advantage (because the onset of acoustic energy was 9 ms earlier for complex words than for simple words), this effect was cancelled by the 20% of triggers that were delayed until the onset of voicing. The combination of these conflicting patterns explains the null result that was obtained in statistical analysis of data from this voice key.

Because the pattern of results obtained appears to depend on the likelihood of the voice key detecting events during the onset segment of each token, it is of interest to determine which acoustic properties of the speech signal predict whether voice keys are triggered during that segment. With this goal in mind, we measured for each token the maximum root mean square (RMS) amplitude of the prevocalic segment, using ESPS/Waves+ (rectangular, 20-ms window; step size = 5 ms). For both voice keys, this measure was greater for tokens detected before the onset of voicing than for those detected at or after that acoustic marker: simple threshold key (613 vs. 271, respectively), $t_1(20) = 5.45$, $p < .01$, and $t_2(39) = 15.48$, $p < .01$;[6] integrator key (417 vs. 334, respectively), $t_1(15) = 2.45$, $p < .05$, and $t_2(39) = 4.36$, $p < .01$.[7] These differences illustrate the importance of acoustic amplitude in voice key performance; factors that influence amplitude, such as individual participant characteristics or fatigue during the experiment, are therefore important variables to consider whenever voice keys are used to measure response latency. Amplitude differences of this type do not influence hand-coded naming latencies as long as a favorable signal-to-noise ratio is achieved, for example, by using a good quality microphone and recording in a sound-attenuated area.

Discussion

From measurements of response latency made using electronic voice keys, experimental psychologists make inferences about the nature and temporal character of the cognitive processes that precede the response. The two aims of this experiment were (a) to investigate voice key error for a common English onset phoneme, using two types of voice key, and (b) to investigate the assumption that voice key error can be controlled by matching initial phoneme across the experimental manipulation, using the onset complexity effect as an illustration. To this end, we compared naming latencies (measured by visual inspection of the speech waveform and by two types of modern voice keys used in psychology laboratories) collected from a single set of participants who read aloud words with simple and complex onsets beginning with the pho-

[6] For 1 participant, the simple threshold voice key never triggered during the onset; hence, this analysis was carried out over only 21 out of 22 participants.

[7] For 6 participants, the integrator voice key always triggered during the onset; hence, this analysis was carried out over only 16 out of 22 participants.

neme /s/. The three measurement techniques produced the three logically possible results: Whereas hand-coding of the acoustic onset revealed a significant complexity advantage similar to that reported by Kawamoto and Kello (1999), the simple threshold key revealed a significant complexity disadvantage similar to that reported by Frederiksen and Kroll (1976), and the integrator key revealed no effect.

With respect to the first of our aims, these results corroborate previous suggestions (Pechmann et al., 1989; Sakuma et al., 1997) that electronic voice keys do not reliably detect the acoustic onset of a syllable. Indeed, in this examination of voice key detection for the /s/ initial phoneme, we found a range of error similar to that reported in these previous studies. As we expected, voice key error was not equivalent across two different types of voice key, however; whereas our simple threshold voice key generally failed to detect any part of the onset segment, triggering instead at the onset of voicing, our integrator voice key generally triggered during the onset segment, on average approximately 80 ms after the acoustic onset.

Given the extent of the voice key error that we have reported, one might ask whether, with better calibration, we might have improved the performance of our voice keys. It is certainly possible that by increasing the sensitivity of our voice keys, we would have reduced the average interval between acoustic onset and voice key detection point. However, we would argue that this decreased level of error in detecting acoustic onset would be accompanied by a catastrophic increase in the number of false alarms (detections before the acoustic onset). This trade-off between accurate onset detection and false alarms is illustrated in Figure 2.

In both of these speech tokens, the amplitude of the onset of the initial phoneme is less than the amplitude of the nonspeech sound (a lip pop in the case of *sect* or an exhale in the case of *staff*) that precedes it. Thus, if the sensitivity of a voice key were calibrated such that it was able to detect the onset of the initial phoneme, the result would be detection of the nonspeech sound for both of these tokens. By raising the amplitude threshold of the voice key, one could avoid detection of the nonspeech sound; however, the consequence would be detection of the speech sound some milliseconds after the acoustic onset.

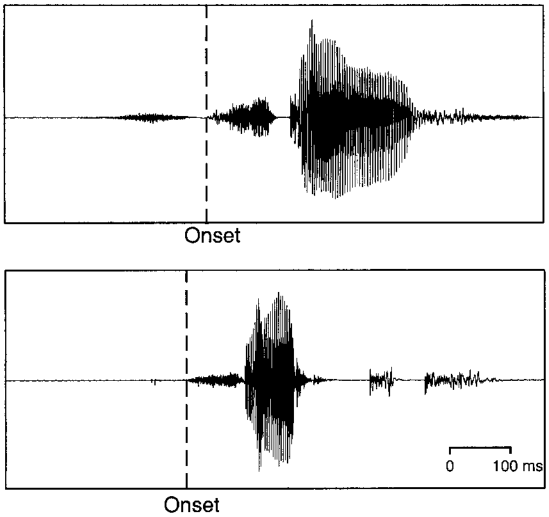

Figure 2. Sample acoustic waveforms for *staff* (top) and *sect* (bottom). In both cases, the acoustic onset of the word (the desired detection point) is preceded by a nonspeech sound that is greater in amplitude than the acoustic onset.

We briefly examined the extent to which this trade-off would be revealed in our data set by conducting a second mock experiment with the integrator voice key. With the knowledge that this voice key had performed poorly in the main experiment, we deliberately increased its sensitivity during calibration. The results of this mock experiment showed a substantial improvement in the detection of acoustic onset; the average detection point was just 45 ms after the true acoustic onset (rather than the 82 ms previously found). Accompanying this decrease in error, however, was a fivefold increase in the percentage of false alarms, from 1.6% of the tokens to 8.5% of the tokens.

Given these figures, it is not clear whether our recalibration of the integrator voice key did, in fact, improve its performance. One might argue that a loss of nearly 10% of the data points is not sufficiently substantial to be of concern; indeed, it would not be surprising to lose 10% of data because of participant errors. However, the point we wish to make here is that without some reference to the acoustic waveform—whether on-line during the experiment or off-line as done here—it is not possible to accurately discard those trials in which the voice key triggered because of some nonspeech sound. Our survey of the articles published in the *Journal* that used a voice key revealed that although most investigators reported discarding trials spoiled by voice key failure, only 3% (one study) identified spoiled trials by referring to the acoustic waveform. Other investigators reported that false triggers were eliminated by data cleaning procedures. However, because only 18% of the false alarm reaction times detected in the mock experiment reported above fell outside of the second standard deviation, these values would be unlikely to be removed by typical data cleaning procedures. Most investigators, we suspect, detected false alarms on-line via a perceptual judgment of the synchronization of the heard acoustic onset and a signal denoting voice key detection, a method whose accuracy and potential bias is clearly questionable. We would argue, therefore, that attempting to decrease the error of a voice key by deliberately increasing its sensitivity (as we have done here) is likely to result in the inclusion of substantial data that may not be indicative of any cognitive process.

Perhaps the most important result that arises from this consideration of the acoustic events detected by voice keys is the challenge we have posed to the widely held assumption that measurement error can be controlled by matching conditions on initial phoneme. Analyses of the speech tokens produced in this experiment revealed significant acoustic differences between simple and complex words matched on initial phoneme, which had a marked effect on the latencies measured by each of the voice keys. For instance, although the initial /s/ was of a shorter duration in the complex context than in the simple context, the total duration of complex onsets (the interval between acoustic onset and onset of voicing) was far greater than that for simple onsets. Because a proportion of the voice key detections occurred at the onset of voicing (which was 31 ms later for complex words than for simple words), differential amounts of voice key error were introduced for simple and complex words. In this instance, these differences (amounting to a maximum 20-ms swing in the mean effect size) were enough to significantly influence the direction of the experimental effect. These results indicate that matching on initial phoneme does not control voice key error across conditions. Instead, we propose that the type of voice key error documented here (that which is created by a voice key that detects some event between the acoustic onset and the onset of voicing) can be adequately controlled only by matching conditions on the complete syllabic onset (i.e., all phonological segments preceding the vowel).

In fact, the need to match experimental conditions on complete onsets goes well beyond the problems encountered when attempting to measure acoustic onsets automatically with voice key hardware. Consider one of the findings reported here—that acoustic energy for /s/ occurred not only later (9 ms), but also over a longer duration (28 ms longer) in the simple context than in the complex context. This effect of segment duration is a well-known phenomenon that is relevant not only to /s/, but to all consonantal onsets, and has been explained in two ways: (a) the shortening of segments in complex clusters is an articulatory phenomenon termed *compression* (Lindblom & Rapp, 1973), and (b) the lengthening of segments in simple clusters is due to the processing requirements of computing vowel phonology. (This second account assumes that speaking begins before all phonemes in a syllable are computed; Kawamoto & Kello, 1999; but see Rastle et al., 2000.) These segment duration effects proved problematic in our experiment when the voice keys failed to detect the onset of acoustic energy. However, acoustic measurements carried out by hand can also be contaminated by differences in initial segment durations. Recall that the production of stop consonants (/k/, /g/, /t/, /d/, /p/, and /b/) consists, in part, of a period during which the vocal tract is completely occluded and no acoustic energy is produced. For words beginning with such consonants, any factor (perhaps onset complexity) that influences the duration of this silent closure period will influence the timing with which acoustic energy is emitted. Only matching syllabic onsets can control the influence of initial segment duration in psycholinguistic experiments.

We have argued that substantial difficulties can be encountered when naming latencies are compared for words that do not have matched onsets. Voice keys may fail to detect initial segments and may thus produce measures of acoustic onset that include the duration of the initial segment (and in some cases the following segment) itself. When these segment durations differ across conditions, sizeable shifts in the magnitude and direction of experimental effects may emerge. Differences in initial segment duration across conditions can also be problematic, even when more reliable means of detecting acoustic onset (e.g., hand-marking) are used, namely for those words that begin with plosive phonemes (because here again, naming latency includes the duration of part of the initial segment). Although differences in initial segment durations across words with unmatched onsets may be small relative to the total time taken to read a word aloud or to name a picture, the size of these differences must be viewed relative to the size of an experimental effect. In the present experiment, we observed a 20-ms shift in the size of the effect due to the interaction between measurement error and complexity condition, which was large enough to change the direction of the effect. The extent to which this degree of error could affect other results similarly would, of course, depend on the size of those experimental effects.

In summary, we have highlighted some of the pitfalls involved in taking acoustic measurements for the purpose of making inferences about cognitive processing. Given our findings, we would suggest to researchers that if a voice key must be used to detect acoustic onset, it should be used within an experimental context in which complete word onsets are matched, to avoid differential degrees of error across the experimental manipulation. We would also argue that more accurate measurements of acoustic onset can be derived by visual inspection of the acoustic waveform (or by using an algorithmic equivalent of this procedure; Kello & Kawamoto, 1998). As explained, however, even these more accurate measures of acoustic onset are not immune to the problems encountered in drawing acoustic comparisons between syllables, and as a result, such comparisons also require matching on complete onsets.

References

Boder, D. P. (1933). Some new electronic devices for the psychological laboratory. *American Journal of Psychology, 45,* 145–147.

Boder, D. P. (1940). A new apparatus for voice control of electric timers. *Journal of Experimental Psychology, 26,* 241–247.

Dunlap, K. (1913). Apparatus for association timing. *Psychological Review, 20,* 250–253.

Dunlap, K. (1921). An improvement in voice keys. *Journal of Experimental Psychology, 4,* 244–246.

Eriksen, C. W., Pollack, M. D., & Montague, W. E. (1970). Implicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology, 84,* 502–507.

Fletcher, J. M., & Bosch, W. C. (1938). A suggested improvement in voice key construction. *Journal of Experimental Psychology, 22,* 97–100.

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

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

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.

Guttman, H. E. (1957). A voice or sound key. *American Journal of Psychology, 70,* 456–457.

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.

Kahn, L. D. (1935). An electronic voice key. *Journal of General Psychology, 12,* 447–450.

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

Kello, C. T., Plaut, D. C., & 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. (in press). Voice key artifacts in vocal response time measurements. *Journal of Memory and Language.*

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

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

Pechmann, T., Reetz, H., & Zerbst, D. (1989). Kritik einer me-smethode: Zur ungenauigkeit von voice-key messungen [The unreliability of voice-key measurements]. *Sprache & Kognition, 8,* 65–71.

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.

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (Eds.), *Speech coding and synthesis* (pp. 495–518). New York: Elsevier.

Vaughn, J., & Strobel, E. J. (1940). A carbon contact voice key. *Journal of General Psychology, 23,* 441–442.

Wells, F. L., & Rooney, J. S. (1922). A simple voice key. *Journal of Experimental Psychology, 5,* 419–427.

Received October 18, 2000
Revision received April 27, 2001
Accepted July 11, 2001 ■