REACTION TIME IN PHONEME MONITORING VARIES WITH SEGMENT DURATION

Paula West, Andrew Slater, John Coleman, and Mario Cortina Borja

Abstract

This paper describes three phoneme monitoring experiments which show that reaction time increases with phoneme duration. Previous studies have shown that reaction time differs according to phoneme type (van Ooijen, Cutler and Norris 1992, Cutler, van Ooijen, Norris and Sanchez-Casas 1996). Our first experiment confirms this finding. Our results also suggested that differences in reaction time are partly dependent on segment duration. Two further phoneme monitoring experiments were conducted to test this hypothesis by manipulating the duration of the target vowel or consonant. In all of the experiments we found that target phoneme duration was a significant factor in reaction times.

1. Introduction

In a phoneme monitoring task, subjects listen to lists of words or non-words and are required to press a button as soon as they hear a stimulus that contains the target phoneme. Phoneme monitoring studies have been used to address a wide range of issues in speech perception, including the contribution of the lexicon, the role of prosodic information and the basic units of speech perception (Connine and Titone 1996). However, the technique is not without its problems. Cutler (1981) points out the many pitfalls open to researchers using the phoneme monitoring technique which may lead to artefacts. Amongst these are timing, word frequency and intensity of the signal. As Cutler writes, the choice of materials in a reaction time (RT) experiment is crucially important: RT has been shown to be greater when there is a mismatch between the phonological structure of foils and the target item (McNeill and Lindig 1973). A mismatch between the target and stimulus context can lengthen RTs (Mills 1980), as can artificial shortening or lengthening of vowels preceding the stimulus (Martin 1979). These results show that RT is extremely sensitive to small changes in stimulus, and that the design of RT stimuli should be undertaken with great care. Our research leads us to add another potential source of artefacts to this list: segment duration.

Effects of signal duration on reaction time to acoustic non-speech stimuli have been reported (Luce 1986, 65-68), although the relationship is not straightforward. For short stimuli (ranging from 2 to 100 ms), RT has been found to be a function of stimulus duration, increasing as stimulus duration decreases. For longer stimuli (100 ms to 2400 ms), RT increases with stimulus duration (Luce 1986). It has been suggested that this may be a result of subjects' using the opportunity to take a longer sample of the sensory information than is necessary (Brebner and Welford 1980).

The results obtained so far using speech stimuli are complicated, partly due to the fact that materials used differ so widely. For vowels in non word-initial position, a clear effect of duration has been reported: longer vowels are responded to more rapidly than shorter vowels (Cutler, van Ooijen, Norris and Sanchez-Casas 1996). However, this effect was found to hold only for medial and not final vowels in another set of experiments (van Ooijen 1994).

Different types of segments have inherently different durations, e.g. stops are typically shorter than fricatives. Thus one might expect different RTs for different segment types, and indeed the effect of segment type on response time has been documented. RTs for vowels have been found significantly longer than RTs for stop consonants, except in initial position. This distinction cannot be simply one of vowel vs. consonant. The semi-vowels /j/ and /w/ (duration 27 to 111 ms) produced longer RTs (mean 628 ms) than the vowels /i/ and /u/ (durations of 108 to 329 ms, mean RT 521 ms) in initial and medial position (van Ooijen, Cutler and Norris 1992). Fricatives have been reported to produce slower RTs than stops (e.g. Rubin, Turvey and van Gelder 1976), yet another study found no significant difference between RTs for fricatives, stops, nasals and glides (Pitt and Samuel 1990). The complicated and sometimes conflicting results of these studies point to the need for an investigation of the role of segment type and duration in phoneme monitoring.

2. Experiment One

2.1 Aim

This experiment was designed to confirm whether reaction time differs when monitoring for consonants rather than vowels. Two consonantal targets, /t/ and /ʃ/, and the vocalic target /ɜ/ were used. From the literature we would expect shorter RTs for the consonants /t/ and /ʃ/ than for the vowel /ɜ/. We also included the whispered vowel [ɜ̥], to test whether RT differences can be attributed to acoustic differences rather than phoneme class.

2.2 Design and stimuli

The experiment was divided into four tasks: monitoring for /t/, /ʃ/, /ɜ/ and [ɜ̥]. The target phonemes were embedded in the following syllables: [tɜ, ʃɜ, ɜʃ, ɜ̥ʃ]. For the consonant monitoring tasks, 8 CV nonsense syllables were prepared as foils: [dɜ, gɜ, tʃɜ, dʒɜ, vɜ, zɜ, lɜ, nɜ]. For the /ɜ/ task, the foils were [iʃ, ɛʃ, ɒʃ, ɔʃ, ʊʃ, uʃ, aɪʃ, aʊʃ]. The same set of foils was used for [ɜ̥], except that the stimuli were whispered. All the stimuli were non-words, to avoid the confounding effect of word frequency and subjects have been reported to focus attention more narrowly on the target when monitoring in pseudowords than in real words (Pitt and Samuel 1990).

The stimuli were recorded by a phonetically trained female speaker of southern British English using an Audio-Technica AT4031 microphone, Symetrix SX202 pre-amplifier and Sony TCD D10 DAT player. They were digitally transferred to the hard disk of a Silicon Graphics Indy workstation, then converted to 16kHz Microsoft WAV audio files. Audio files containing instructions for the subjects were prepared in the same way.

The duration of stimuli and the target phonemes naturally varied. Table 1 gives the duration of the target phonemes:

Table 1: Duration of targets, experiment 1

Target	t	ʃ	ɜ	ɜ̥
duration (ms)	94	173	236	373

For each of the 4 tasks, 60 randomisations of each set of 9 syllables were combined to create a list of 540 stimuli for presentation to subjects. A practice list with a different target was also prepared.

2.3 Subjects and procedure

The subjects were six right-handed male monolingual speakers of southern British English, aged 18 to 22. They were all students of University of Oxford, without training in linguistics or phonetics, and were paid for their time.

Stimuli and audio instructions were presented over Sennheiser HD320 headphones in a sound-treated room via a standard Soundblaster 16 card installed in a Pentium PC. Subjects were given a response button (a push-to-make switch) which was connected to a timer I/O card (hardware manufactured by technical staff in the Department of Experimental Psychology, Oxford) installed in the same PC. Stimuli were presented and subjects' responses were collected using software developed in-house. Reaction times were measured from the onset of the audio stimulus until the button was pressed. The interstimulus interval was approximately 1200ms.

Before each run, subjects were told the target sound and instructed to press the button as fast as possible each time they heard it. The experiment was divided into a practice session and four experimental runs, one per target. The order of presentation of the four tasks was randomised between subjects. Subjects were given a short training session immediately prior to the experiment, to familiarise them with the procedure.

2.4 Results and discussion

Table 2 gives mean reaction times for each target averaged across subjects. Contrary to our hypothesis, we found no clear division between responses to consonant and vowel targets. Rather, mean RTs for /t/ were approximately 80ms shorter than for /ʃ/, /ɜ/, [ɜ̥]. Shorter RTs for stop consonants in real words have been reported by van Ooijen, Cutler and Norris (1992), although not in word-initial position.

Table 2: Mean RT, experiment 1

Target	t	ʃ	ɜ	ɜ̥
mean RT (ms)	407	490	467	501

The reaction times were log transformed to reduce the effect of outliers, and a one-way repeated measures analysis of variance was constructed in SPLUS with target phoneme as the independent variable and log RT as the dependent variable. There was a highly significant effect ($F(3,1340) = 39.27$, $p < 0.001$) for target type.

Although our results confirm that RTs differ according to phoneme, our failure to find the expected difference between consonants and vowels led us to consider confounding factors. One such factor might be duration. The durations of the target phonemes are plotted against their mean RTs in Figure 1.

Figure 1 Mean RT against target duration, experiment 1

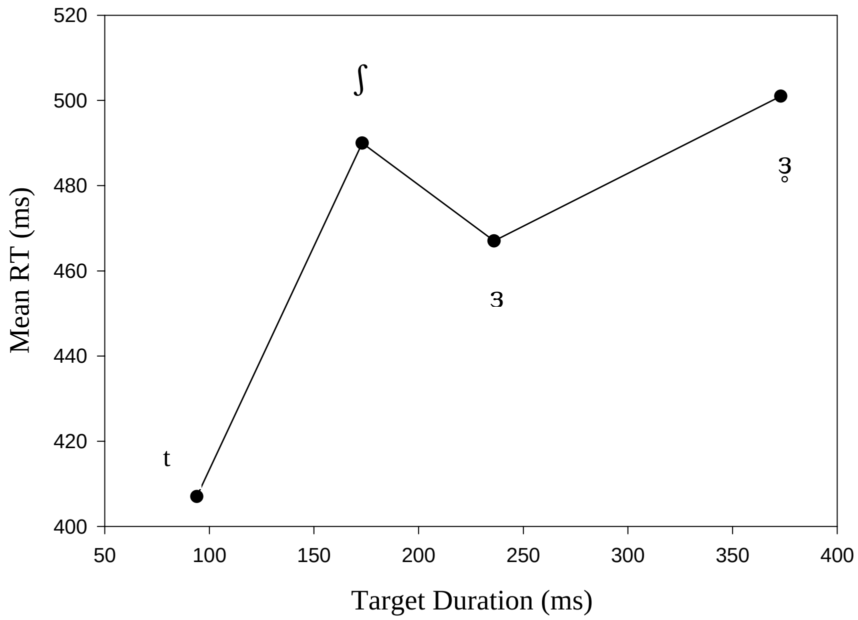


Figure 1 suggests a tendency for RT to increase with target duration. To examine this trend further, a linear mixed effects model for repeated measures was constructed, with target as a linear fixed effect factor, ordered by target duration. Target was highly significant (z ratio = 5.57, $p < 0.01$). This confirms that RT differs for at least two targets. Post-hoc linear mixed effects models constructed for pairs of targets confirmed the trend for RT to increase with duration, as follows:

log RT(t) < log RT(ʃ,ʒ) < log RT(ʒ̊) ($\alpha = 0.01$)

To test the hypothesis that RT varies with target phoneme duration, two further experiments were designed and conducted.

3. Experiment Two

3.1 Aim

This experiment was designed to test whether reaction time in monitoring for a consonant varies with the duration of the target segment. The phoneme chosen was the voiceless fricative /ʃ/ as digital recordings of natural voiceless fricatives can easily be increased or reduced in duration, using simple waveform-editing tools, thus avoiding the difficulty and pitfalls of making high-quality synthetic speech stimuli.

3.2 Design and stimuli

The experimental design was a slightly modified version of that of the first experiment. The task was to monitor for /ʃ/, and the duration of the target and foils was varied. Target [ʃ$_1$] was the naturally-produced /ʃ/ used in experiment 1. [ʃ$_2$], [ʃ$_3$] and [ʃ$_4$] were created by copying and inserting sections of [ʃ$_1$], using the *xwaves* speech processing package (Entropic Research Laboratory Inc., Washington). The duration of the four targets is given in Table 3.

Table 3: Duration of targets, experiment 2

Target	$ʃ_1$	$ʃ_2$	$ʃ_3$	$ʃ_4$
duration (ms)	173	230	289	347

Five foils were used: /dʒɜ, vɜ, zɜ, lɜ, nɜ/, with initial consonants of 4 different durations. (/dɜ/ and /gɜ/, which were used in experiment 1, were excluded as altering the duration of stop consonants is impractical. /tʃ/ was excluded as a short /ʃ/ and /tʃ/ are easily confused.) The durations of the initial consonants were adjusted in the same way as the target, by inserting and/or removing sections of the naturally-produced consonants. The duration of the vowel /ɜ/ in the targets and foils was not altered.

60 randomisations of each set of 6 syllables were combined to create four lists of 360 stimuli, presented as four separate runs. Each of the four runs contained a target and foils of the appropriate relative duration.

3.3 Subjects and procedure

Experimental procedure was as described for experiment one. Five of the subjects from the first experiment were recalled for this experiment. The sixth subject was not recalled, as his data had shown a different pattern of response times from the other 5 subjects.

3.4 Results and discussion

Mean reaction times for each target averaged across subjects are given in Table 4 and are plotted against target duration in Figure 2.

Table 4: Mean RT, experiment 2

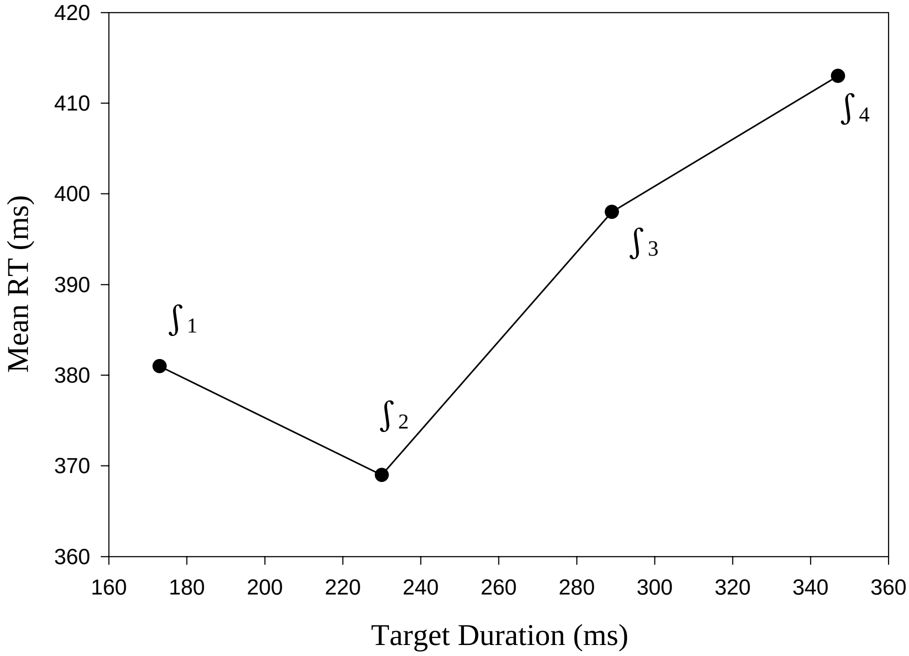
Target	$ʃ_1$	$ʃ_2$	$ʃ_3$	$ʃ_4$
RT (ms)	381	369	398	413

As for experiment 1, the reaction times were log transformed to reduce the effect of outliers, and a one-way repeated measures analysis of variance was constructed in SPLUS with target as the independent variable and log RT as the dependent variable. There was a highly significant effect ($F(3,1183) = 13.15$, $p < 0.001$) for target.

A linear mixed effects model for repeated measures was constructed, with target as a linear fixed effect factor, ordered by target duration. Target was highly significant (z ratio = 2.31, $p = 0.0104$). This confirms that RT differs for at least two targets. Post-hoc linear mixed effects models constructed for pairs of targets confirmed the trend for RT to increase with duration, as follows:

$$\log RT(ʃ_{1,}ʃ_2) < \log RT(ʃ_{3,}ʃ_4) \quad (\alpha = 0.01)$$

Figure 2 Mean RT against target duration, experiment 2



The results of this experiment strongly suggest to us that (at least for fricatives) reaction times in phoneme monitoring tasks vary according to the duration of the target, as suggested by the results of Experiment 1. There is no way of easily changing the duration of a plosive (at least in stimulus-initial position), and so we draw our conclusions about consonants on the basis of those whose duration can be easily manipulated.

However, there is another possible interpretation of this result. The effect of duration may be due to the grouping of targets and foils of similar relative duration within a run. If all stimuli are of relatively short duration, responses to targets might be faster. The experimental design does not allow us to decide whether the result is due to the blocked design or to target duration.

4. Experiment Three

4.1 Aim

Experiment three was designed to explore the effect of duration further, avoiding the confounding factor of experiment two. In this experiment, a vocalic target was used.

4.2 Design and stimuli

The targets were three tokens of /i/ of different duration (see Table 5). These were created using the time stretch function in *Cool Edit 96* (Syntrillium Software Corporation). A time stretching algorithm (rather than wave-form concatenation) was needed to preserve the pitch contour and create natural-sounding stimuli.

Table 5: Duration of target phonemes, experiment 3

Target	i_1	i_2	i_3
Duration (ms)	134	222	377

Eight foils were used: /ɜʃ, ɛʃ, ɒʃ, ɔʃ, ʊʃ, uʃ, aɪʃ, aʊʃ/, each with initial vowels of three different durations, created using *Cool Edit 96*. The consonant durations were unaltered. The resulting set of 24 foils and 3 targets were randomised in 20 blocks to create lists of 540 stimuli. Five such lists were prepared, presented to subjects as five separate runs.

4.3 Subjects and procedure

The six subjects for this experiment came from the same population as those of the earlier experiments. The procedure was the same as that followed in experiments 1 and 2.

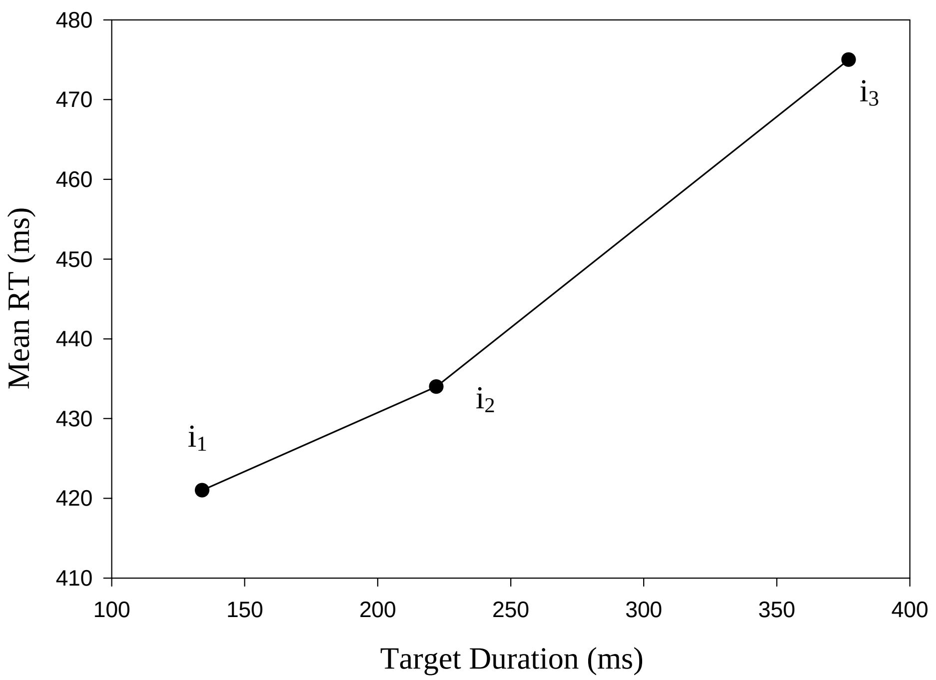
4.4 Results and discussion

Mean reaction times for each target averaged across subjects are given in Table 6 and are plotted against target duration in Figure 3.

Table 6: Mean RT, experiment 3

Target	i_1	i_2	i_3
Mean RT (ms)	421	434	475

Figure 3 Mean RT against target duration, experiment 3



As for experiments 1 and 2, the reaction times were log transformed to reduce the effect of outliers, and a one-way repeated measures analysis of variance was constructed in SPLUS with target as the independent variable and log RT as the dependent variable. There was a highly significant effect ($F(2,1724) = 36.54$, $p<0.001$) for target.

A linear mixed effects model for repeated measures was constructed, with target as a linear fixed effect factor, ordered by target duration. Target was highly significant (z ratio = 5.73, $p < 0.001$). This confirms that RT differs for at least two targets. Post-hoc linear mixed effects models constructed for pairs of targets confirmed the trend for RT to increase with duration, as follows:

$$\log RT(i_1) < \log RT(i_2) < \log RT(i_3) \quad (\alpha = 0.01)$$

This experiment shows that the duration effect was not simply a result of the blocked design of our earlier experiments, and convinces us that duration may affect reaction time in phoneme monitoring tasks. Given that we did not vary the duration of the final consonant, the target duration and the entire stimulus duration are completely correlated. Therefore, the effect may be due either to target duration or stimulus duration. Further experiments are needed to separate these two factors.

5. Discussion and conclusions

We have shown that duration has a significant effect on reaction times in phoneme monitoring experiments, and found no consistent effect of the vowel-consonant distinction.

Previous studies (e.g. Cutler et al. 1996) which have found differences between reaction times to vowels and consonants suggest that this is due to different processing strategies resulting from greater variability in vowels than consonants.

Our first experiment suggested that differences in reaction times to consonants and vowels might be a result of differences in inherent segment duration. Experiments two and three, which manipulated the duration of a single vowel and consonant target confirm that duration has an effect, although this may not be due to target duration alone. Nevertheless, we have shown that duration is an important factor that must controlled for in phoneme monitoring experiments.

References

Brebner, J.M.T. and A.T. Welford. 1980. Introduction: An Historical Background Sketch, in Welford, A.T. (ed.) *Reaction Times*. London: Academic Press.

Connine, C.M. and D. Titone. 1996. Phoneme Monitoring. *Language and Cognitive Processes* 11(6), 635-645.

Cutler, A. 1981. Making up materials is a confounded nuisance, or: will we be able to run any psycholinguistic experiments at all in 1990? *Cognition* 10, 65-70.

Cutler, A., B. van Ooijen, D. Norris and R. Sanchez-Casas. 1996. Speeded detection of vowels: A cross-linguistic study. *Perception and Psychophysics* 58, 807-822.

Luce, R.D. 1986. *Response Times*. Oxford: OUP.

Martin, J.G. 1979. Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America* 65(5), 1286-1297.

McNeill, D. and K. Lindig. 1973. The perceptual reality of phonemes, syllables, words and sentences. *Journal of Verbal Learning and Verbal Behaviour* 12, 419-430.

Mills, C.B. 1980. Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behaviour* 19, 75-83.
Rubin, P., M.T. Turvey and P. van Gelder. 1976. Initial phonemes are detected faster in spoken words than in non-words. *Perception and Psychophyics* 19, 394-498.
van Ooijen, B., A. Cutler and D. Norris. 1992. Detection of vowel and consonants with minimal acoustic variation. *Speech Communication* 11, 101-108.
van Ooijen, B. 1994. *The processing of vowels and consonants*. Doctoral dissertation, University of Leiden.

Contact:

Paula West, Andrew Slater, John Coleman
Oxford University Phonetics Laboratory

Mario Cortina Borja
Oxford University Department of Statistics