A Syllabic Bottleneck in Prelexical Processing ?
A Phoneme Monitoring Investigation

E. Dupoux
France Telecom, Paris; EHESS–CNRS, Paris

LSCP TECH REPORT 94-2
September, 13th, 1994

Previous research has found that phoneme detection latencies depend on the complexity of the syllable that bears the target phoneme. CV syllables give rise to faster latencies than CVC, that are faster than CCV (Treiman *et al.*, 1982, Cutler *et al.*, 1987). In Experiment 1, we replicate this result and extend it to a fourth structure: CCVC. In Experiment 2, we report a similar effect in first syllables of disyllabic items, showing that complexity effects cannot be reduced to stimulus duration effects. We argue that the complexity effect is inconsistent with the view that phonemes are the only units involved in speech perception, but supports models which stipulate larger sized units like syllables (Mehler, 1981; Segui, Dupoux & Mehler, 1990). In a series of post-hoc analyses, however, we show that the complexity effect is not uniform across subjects. Although both the complexity of onsets and codas of syllables influence phoneme detection latencies for slow subjects, fast subjects are only influenced by the nature of the onset. The interaction of speed of response with complexity effects is confirmed in Experiment 3, where it is found that when subjects are urged to respond as fast as possible, CVC items no longer show a complexity effect nor a lexical superiority effect. Implications for the existence of a syllabic bottleneck and the time course of prelexical processing are discussed.

The quest for the basic unit involved in speech processing is not a new one. Proposals ranging from distinctive units (Eimas & Corbit, 1973) to syllables (e.g. Savin & Bever, 1970), to entire words (Klatt, 1980) have been put forward. Nowadays, there is fairly general agreement that the basic linguistic unit involved in processing is the phoneme (Marslen-Wilson, 1984; McClelland & Elman, 1986; Pisoni & Luce, 1987). Congruent with this, is the assumption that speech perception must proceed in a smooth and continuous fashion. Hence the units involved should be as small as possible to allow for lexical access to occur as fast as possible (Warren & Marslen-Wilson, 1987; 1988). For a number of researchers, however, such a requirement of optimal efficiency does not hold. Instead, they think that the processing system is better off if it comprises discrete stages, whereby at each stage, information is accumulated until a stable representation is achieved,before being passed on to the next stage (e.g. Sternberg, 1969). In particular, it has been proposed that speech processing relies on the accumulation of information in windows about the size of a syllable (Savin & Bever, 1970; Mehler, 1981; Mehler, Dupoux & Segui, 1990; Segui, Dupoux, & Mehler, 1990). In this view, phonemes are extracted only after syllable-like representations have been stabilized.

There are two sides to this issue: one is related to the size and format of the processing units (e.g. syllables versus phonemes). The other side concerns the question of whether processing is continuous, or whether there exists discrete stages at which processing stops until enough information has accumulated. These two issues are related since proponents of larger-sized units have tended to view processing in a discontinous way, whereas believers in smaller sized units have favoured continuous processing. However, to a certain extent, the two problems can be dissociated: one could have syllables as basic units, but let partial activation of multiple syllable units 'cascade' through the system. Conversely, one could imagine a system based on phonemes, but incorporating a discrete stage whereby three or four phonemes would be accumulated before triggering, say, lexical access. In this paper, we will address these two issues (type of unit and con-

---

This paper was written while the author was a visiting researcher at the Psycholinguistics lab at the University of Arizona, Tucson, and at the Institute for Research in Cognitive Science, at the University of Pennsylvania, Philadelphia. Part of this work was presented in two book chapters: Mehler *et al.*(1990), and Dupoux (1994). I would like to thank A. Christophe, L. Gleitman, J. Mehler, J. Miller, C. Pallier, B. van Oyen, and J. Segui, for their extremely useful comments and helpful discussion. I would also like to thank C. Bonneuil and M. Binagi, who helped to run some of the experiments reported here. This research was supported in part by CNET (Convention 837BD28 00790 9245 LAA/TSS/CMC), CNRS (ATP "Aspects Cognitifs et Neurobiologiques du Langage"), the European Science Foundation (TW 86/17) and the Ministere de la Recherche et de l'Enseignement Superieur decision nb 84C1390, INRIA, the Fyssen Foundation and the Human Frontier Science Program.

Address for correspondence: LSCP CNRS-EHESS, 54 Bd Raspail 75006 Paris FRANCE; email: dupoux@lscp.ehess.fr, web: http://www.ehess.fr/centres/lscp/persons/dupoux.htm

tinuity of information flow) by exploring how the detection of individual phonemes is affected by the syllabic structure in which they appear.

The claim that syllable-sized units are involved somewhere in the processing system has been supported by an extensive body of data, at least for romance languages like French and Spanish (Mehler *et al.*, 1981; Dupoux & Mehler, 1990; Sebastian *et al.*, 1993; Pallier *et al.*, 1993). For these languages, a view by which only phonemes are used in speech processing cannot account for the data. What is still controversial is whether syllabic units are the 'first' ones or whether they are derived from infra-syllabic units. A more simple basic question which has not be addressed yet, is whether these syllabic units, wherever they are located in the system, act as information bottlenecks, that is, whether they only let information percolate through once the whole syllable has been processed.

But what evidence do we have that there are any bottlenecks at all? First, experiments cited in support of the continuous and incremental view of speech processing are in fact equivocal. This can be illustrated with gating studies relying on fragments cut out from the signal (see Warren & Marslen-Wilson, 1987; 1988). These studies show that the processing system makes use of all acoustic information available in a given signal fragment whatever its size. However, this does not disprove the existence of bottlenecks since the end of the fragment might itself be interpreted as the end of the processing unit, and trigger the release of the information accumulated so far. Second, many experiments report that phonetic information relevant to a given segment is integrated over a comparatively large processing window. For instance, the acoustic cues to the identity of a given consonant in a CV syllable depend, among other things, on the quality of the following vowel (e.g. Bailey & Summerfield, 1980; Repp, Liberman, Eccardt & Pesetsky, 1978) and also on its physical duration (Miller & Liberman, 1979; Miller, 1981). Such supra-phonemic effects are also evident in speeded phoneme detection tasks. Reaction times (RTs) to the first phoneme of a CV sequence are correlated with the vowels' typical physical duration (Foss & Gernsbacher, 1983; Diehl, Kluender, Foss, Parker & Gernsbacher, 1987). This means that in order to determine what the first consonant is, the system does take into account certain characteristics of adjacent segments (see also Fowler, 1984). One reason for this might lie in coarticulation, that is the fact that acoustic information concerning consonants and vowels do overlap. Alternatively, it might be that even when acoustic information does not overlap, the processing system still takes into consideration more than one segment at a time and perceives them, so to speak, in an integral fashion.

One example of this latter case is when the duration of the vowel in synthetic CV syllables affects the categorization of the initial consonant. Even though the initial transition (which should be sufficient to identify the first consonant) remains constant, the perceptual system takes into account the whole CV stimulus (Miller & Liberman, 1979). Also, when subjects have to classify the first segment of CV stimuli, they are slowed down when the following vowel varies in identity (e.g. Wood & Day, 1975; Tomiak *et al.*1987). Importantly, such integral perception effects arise even in case of synthetic stimuli in which coarticulation is absent, and where the identification of C *could* have been made independantly of the V (Tomiak *et al.* (1987).

Finally, there is a more theoretical reason for seriously considering processing bottlenecks in speech recognition. Because of coarticulation, a system updating its representation every millisecond would yield many local and incorrect guesses about the current phonemic composition of the signal. These incorrect guesses might then cascade and activate spurious lexical candidates that would compete with the real candidate, only to be discarded by later occurring information. In other words, continuous uptake of information does not yield obvious processing advantages, but also potentially increases the amount of noise in the system, making speech perception more difficult. In contrast, a system whereby information about a larger chunk of speech input is accumulated and stabilized before being passed on to further stages makes good processing sense, only at the relatively small expense of a delay in information transmission (see Mehler *et al.*1990 for a similar argument).

In brief, the idea that the information processing grain is larger than the phoneme has both empirical support and theoretical appeal. The issue we want to address in this paper is whether the syllabic unit whose relevance for processing has been demonstrated in various studies, is a bottleneck in processing or not. Mehler *et al.*have proposed a model according to which the answer to this question would be yes. They posited a system of syllabic detectors (syllabogens) that would only fire once a syllable has been recognized. Alternatively, one might incorporate syllabic units in a model such as TRACE (McClelland & Elman, 1986) as an intermediate level between phonemes and words. In this case, syllables would not function as a bottleneck as information would continuously flow through this level.

In order to address these issues, we wanted to examine in more detail some findings suggesting that the detection of a phoneme in a stimulus depends on durational or structural properties of the entire syllable in which it appears. Segui, Frauenfelder & Mehler (1981) reported that phoneme detection times correlate with syllable detection times. If true, such correlation, together with the claim that syllables are detected faster than phoneme (Savin & Bever, 1970), might indicate that identifying phonemes depends on the prior identification of the syllable. However, the comparison phoneme versus syllable detection time is not a straigtforward one (see Dupoux, 1993 for a discussion). Moreover, in the Segui *et al.*(1981) study, it is not clear whether the observed correlation is due to the acoustic properties of the first segment only or whether there is also an effect of properties of the entire syllable. Other studies, however, found that the latency to detect a phoneme seems to be related to the complexity of the syllable in which it appeared. Treiman, Salasoo, Slowiaczek & Pisoni (1982) reported that the first consonant in a CV syllable was detected faster than that in a CCV syllable. They also found that CV syllables gave marginally faster responses than CVCs (see also Cutler, Mehler, Norris

& Segui, 1987). This result seems to imply that the detection of a given phoneme is in some way or another dependent on the prior processing of the entire syllable, that is, a syllabic bottleneck affects the identification of phonemes.

Yet the effect of syllabic complexity have not received adequate empirical attention. The first paper that reported them remains unpublished (Treiman *et al.*, 1982), and some of the other observations were made with materials that were not specifically designed to address this issue (Cutler *et al.*, 1987). Given the potential importance of the effect of syllabic complexity for the issue of processing bottlenecks, we decided to first replicate this observation in better controlled experiments. Experiment 1 tests the existence of complexity effects in monosyllabic items, and Experiment 2 extends it to disyllabic items. The remainder of the paper examines in more detail the possibility of a syllabic processing bottleneck.

Experiment 1The Complexity Effect

In the following experiment, we studied initial phoneme detection times in monosyllabic items of four different structures: CV, CVC, CCV and CCVC. One can view the choice of these four types of syllables as the result of the crossing of two structural variables: a) simple versus complex onset and b) presence or absence of coda. Extrapolating from the experiments reported above, one would expect the following ranking in reaction times: CV<CCV and CVC<CCVC because syllables with simple onsets should be detected faster than syllables with complex onsets, and CV<CVC plus CCV<CCVC because syllables with no coda should yield faster RTs than syllables with a coda. Furthermore, the effects of onset and coda should be additive. In this experiment, we made sure that subjects were not biased, that is, expect one structure rather than the other. Thus, each target phoneme was presented with a set of four examples, one for each structure, and each of the four structures occured equally often.

Method

Materials.

Eight quadruplets of CV, CVC, CCV and CCVC monosyllabic items were selected. All monosyllables in a given quadruplet shared the vocalic nucleus and the initial consonant (e.g.: /pɑ/,/pɑd/,/prɑ/,plɑd/) and were legal French syllables. The initial consonants were either voiced (/b/, /d/) or unvoiced, (/p/, /t/). A given initial consonant was used with one of two vowels (e.g. /ɑ/ or /o/). This resulted in 32 monosyllabic items (see Appendix) which were mostly non-words. An additional set of 32 fillers was added; these had the same distribution of syllabic structure as the experimental targets. All targets starting with the same consonant were pooled together in a single uninterrupted list, to which an equal number of fillers added. This resulted in 4 different lists, where half of the items were experimental and half were distractors.

The lists were recorded by a native speaker of French at a rate of one word every two seconds. The materials were digitized, and by means of a waveform editor inaudible marks were placed at the onset of the burst of the first consonant in each experimental item. These marks served to trigger a clock and record subjects' reaction times.

Procedure. Before each experimental list, subjects were presented with the target phoneme which was specified auditorily with four non- experimental items of the form CV, CVC, CCV and CCVC (e.g., for the phoneme target /p/: "the target is /p/ as in PLI, PE, PRU, PRED"; the order of the syllabic structuers was randomly varied from one trial to the next). Before starting the experiment proper, subjects were given a warm-up list constructed along the same lines as the experimental lists but with the target phoneme /g/.

Subjects. Two groups of 20 subjects (students at the University of Paris V) received course credit for participating in the experiment. The first group received the two lists with the voiced phoneme targets /b/ or /d/, the second group heard the non-voiced phoneme targets /p/ or /t/.

Results

After running the experiment, it turned out that one item in the /p/-/t/ series, ”pro”, had not been marked and hence no reaction time had been collected. We discarded this item from the analysis, as well as the three other matched items ”po”, ”pob”, and ”prot”.[1]

The mean RTs (ms) of this experiment are displayed in Table 1. Two ANOVAs were run, one with subjects, and another with items as random variables. The ANOVAS contained two within-subjects factors, Onset Complexity and Coda Complexity, and one between subjects factor, Voicing. Both complexity factors yielded a significant effect. Syllables with simple onsets were responded to faster than syllables starting with a consonant cluster ($F_1(1,38) = 31.03, p < .001; F_2(1,6) = 20.4, p < .004; min\ F\ (1,15) = 12.3, p < .003$), and open syllables yielded faster responses than closed syllables ($F_2(1,38) = 8.08, p < .008; F_2(1,6) = 9.88, p < .02; min\ F\ (1,24) = 4.44, p < .05$). The two factors did not interact with one another ($F < 1$) nor did they interact with the voicing of the target phoneme ($F < 1$).

Post-hoc contrasts showed that targets in CV syllables yielded faster reaction times than targets in any of the other structures ($p < .03$ at least). CCVC syllables yielded the slowest reaction times; all comparisons were significant ($p < .003$), except for the one between CCVC and CCV, which was only marginally significant ($.1 > p > .05$). CVC did not differ significantly from CCV ($p > .1$).

Discussion

The results obtained in this experiment clearly indicate a close relationship between initial phoneme detection times and the complexity of the target-bearing syllable. Both the presence of an initial consonant cluster and the presence of a final consonant significantly slow down RTs to initial consonants. Thus, we have replicated and extended the previously

---

[1] Removing or not removing these items did not change the results of the analysis of variance.

Table 1

Mean RTs (msec) to phoneme detection as a function of item target structure in Experiment 1.

	Onset Structure						Onset Complexity Effect
Coda structure	− cluster			+ cluster			
− coda	CV	443	(98)	CCV	492	(136)	49*
+ coda	CVC	472	(105)	CCVC	519	(104)	47*
Coda Effect		29*			27˜		

Note.
Besween parentheses are standard deviations.
(*) $p < .05$
(˜) $.05 < p < .1$

obtained observations by Treiman *et al.*(1982) and by Cutler *et al.*(1987).

Prima facie, these results suggest that in order to recover the identity of the initial phoneme of a stimulus, the perceptual system integrates the acoustic information that spans the entire syllable. If only the first CV (or CCV) portion of the syllable were needed, it is hard to understand why CVC syllables yield slower reaction times than CV syllables. The fact that both onset structure and coda structure affect reaction times suggests that the whole syllable is taken into account. In other words, one could propose a bottleneck whereby information regarding the entire syllable is accumulated until further processing is performed. Only then, as a second step, would phonemes be recovered (Mehler *et al.*, 1990; Segui *et al.*, 1990).

However, this is not the only way to look at the data. The results may have little to do with bottlenecks, but rather reflect task-specific response strategies. For instance, instead of reflecting syllable complexity, reaction times may simply be sensitive to stimulus duration. It could be, for instance, that the end of the stimulus acts as a warning signal for subjects, urging them to respond. In short items (such as CVs) this warning signal would occur comparatively early, whereas in longer syllables (such as CCVC) it would occur late. Thus, the obtained effect could be a length effect that correlates with complexity, but not a complexity effect per se. To test this, one has to use polysyllabic items and manipulate the complexity of the first syllable. In polysyllabic items, since response times usually occur before the end of the stimulus, it seems very unlikely that a stimulus length effect could account for the pattern of results. Experiment 2 was designed in part to address this issue.

Another possible artifact relates to the “*model*” effect discussed by Cutler, Butterfield & Williams (1987). The claim is that in many phoneme monitoring experiments, the target phoneme is specified with a *model*, usually a CV syllable: /b/ as in “boo”. This would induce subjects to match not only the target *phoneme* but also the target *model* with the incoming signal. The more complex the syllable, the greater the mismatch with the memorized *model*, and hence the longer the reaction time. In other words, the syllabic complexity effect would not reflect processing bottlenecks at all, but rather the use of syllabic *models* in decision making. This view predicts that if one were to change the way in which subjects represent the target phoneme, the complexity effect would disappear hold. For instance, if one could force subjects to represent the target with a CVC *model*, then closed syllables should now yield faster RTs than open syllables. Similarly, if one could force subjects to represent the target with a CCV *model*, then syllables with a complex onset should now yield faster RTs than syllables with simple onsets.

Cutler, Butterfield and Williams (1987) made precisely this claim. They found that when the instructions are to detect ‘/b/ as in BLUE’, RTs are shorter for the stimulus “BLEND” than for the stimulus “BESK”.[2] The conclusion they drew is that the *model* given to subjects to specify the phoneme determines the mental representation and hence influences which stimulus will be responded to faster.

Could this type of explanation apply to Experiment 1? The short answer is no. In our experiment, we gave 4 examples for each target (/p/ as in PE, PRO, PID and PLUK), and the order of the examples was random from trial to trial. So it cannot be the case that the *model* we gave to the subjects was responsible for the obtained complexity effect: all four syllable types were presented. One could still argue that when subjects are provided with too many *models*, they only keep the simplest one in memory (CV). In this view, syllabic complexity would have an effect in memory, but not necessarily in processing. That is, subjects would prefer to memorize simple structures rather than complex ones, and complexity effects could be accounted for, in terms of a match-mismatch between the stimulus and the way in which subjects memorize the target phoneme.

In the following experiment, we address the issue of stimulus duration by testing the influence of syllabic complexity in the first syllable of disyllabic items. We also address the potential role of the *model* by using a design similar to that used by Cutler, Butterfield and Williams (1987) with disyllabic words where type of *model* and initial syllable were manipulated independently with three syllabic structures: CV, CVC and CCV.

---

[2] They also found that ‘/b/ as in BLUE’ yielded faster RTs to BLEND than to BREAK, which means that the information regarding the second consonant of the cluster was kept in the subjects’s mental representation of the target phoneme.

Table 2
Mean RTs (msec) to phoneme detection as a function of item target structure and model structure in Experiment 2.

Target structure	Model structure						
	CV		CVC		CCV		Mean RT
CV	341	(112)	378	(104)	389	(115)	369
CVC	373	(122)	390	(126)	397	(108)	393
CCV	420	(103)	414	(103)	409	(97)	414
Mean RT	378		394		398		

Experiment 2

The aim of this experiment is to investigate a) whether syllabic complexity effects can also be observed in polysyllabic words, and b) whether complexity effects can be reduced to *model* effects. In this experiment, subjects had to detect initial phonemes in short lists of disyllabic words. The initial syllable of the target bearing item was either a CV, a CVC or a CCV. As in Cutler, Butterfield and Williams (1987), the target phoneme was specified with a monosyllabic example (the *model*), which was either a CV, a CVC or a CCV word. The structure of the *model* and the structure of the target bearing item were fully counterbalanced in the experimental lists.

The prediction of the *model* hypothesis is that RTs solely depend on the degree of match between the target and the *model*, so irrespective of the complexity of the target. For instance, with CCV *models*, one should find CCV<CV, and CCV<CVC. In case of a perfect match between target and *model*, there should not be any difference in reaction times between CV, CVC and CCV stimuli. In contrast, the prediction of the syllable hypothesis is that a complexity effect should obtain over and above a potential effect of match between target and *model*. For instance, even when target and *model* are perfectly matched, one should still find CV<CVC and CV $<$ CCV.

Method

Materials.

Experimental stimuli consisted of six triplets of French disyllabic words. Each triplet was composed of three words sharing the same initial phoneme, but differing in syllabic structure: CV, CVC or CCV (e.g.: *palace, palmier, plateau*, see the Appendix for a listing). The experimental items were placed in final position of a short sequence (from 1 to 6 items long). There were three experimental blocks, each containing the 18 experimental sequences plus 12 filler sequences. In each block, a particular experimental item was associated with a specific *model*. The *model* was a monosyllabic word corresponding to the initial syllable of a member of a triplet. For example, the experimental sequence bearing the item PALMIER was preceded in one block by: “/p/ as in the word /pal/”, in another block, by “/p/ as in the word /pa/”, and in the third block, by the instruction “/p/ as in the word /pla/”. So in this example, there was an exact match between the *model* and the initial syllable of the experimental item in the first block, but not in the second or third blocks. For other experimental items, the matches or mismatches occured in other blocks such that, globally, there was an equal number of matches and mismatches across the three blocks.

The three blocks were read by a native speaker of French. The pauses between the words in a sequence were 1500 ms long with and 10 seconds of silence between the sequences. Response timing and storage was similar to the previous experiment.

Procedure.

Each subject heard the three experimental blocks. Presentation order was counterbalanced across subjects. Hence, each subject heard, for example, the word “*palace*” three times, each time with a different *model.*

Subjects.

Thirty subjects, Psychology students at the University of Paris V, were tested.

Results

Table 2 shows the mean phoneme detection times (ms) as function of target and *model* structure. An analysis of variance taking subjects as a random factor revealed a highly significant effect of target structure ($F(2,58) = 16.5, p < .0005$), whereas an effect of type of *model* failed to reach significance ($F(2,58) = 3.05, p > .10$). The interaction between the two factors was not significant ($F(4,116) = 2.24, p > .10$). Post-hoc contrasts showed that CCV targets yielded slower responses than both CV and CVC (p<.002), and that CVC targets yield slower responses than CV (p<.03).

To study the role of the syllabic structure of the target bearing item when the effect of the *model* is neutralized, we compared the mean values corresponding to the three cases in which there was a strict match between the *model* and the item structure, i.e.: the three diagonal cells in Table 2. The observed difference was highly significant ($F(2,58) = 11.3, p < .0005$). The ranking of RTs as a function of item structure was the same as in Experiment 1, namely, CV $<$ CVC $<$ CCV. Post-hoc contrasts showed that CV targets yielded faster response times than CVC and CCV (p<.003), but that CVC was not significantly faster than CCV.

Finally, we compared the matching and non-matching responses in order to assess the role of *model*-structure correspondence. The overall difference between matching and non-matching responses was significant ($F(1,29) = 4.19, p < .05$). This agrees with the *model* hypothesis presented in Cutler, Butterfield and Williams (1987). However the match-mismatch effect interacted with item structure

Table 3
Mean RTs (msec) to phoneme detection for matching and nonmatching target-model structure in Experiment 2.

	Target Structure		
	CV	CVC	CCV
Matching	341	390	409
Nonmatching	384	385	417
Effect of match	43*	-5	8

$(F(2,58) = 4.52, p < .02)$ such that this effect mainly applies to CV items (p<.001) but much less so to the other two types of syllabic structure (see Table 3).

Discussion

The results observed in this experiment confirm the influence of syllable structure on phoneme monitoring times. Even in disyllabic items, an effect of syllabic complexity has been uncovered. Given that subjects responded before the end of the stimuli, this means that we can discard duration as the sole factor explaining syllabic complexity effects. Some other factor relating to the structure of the stimulus has to be considered.

Second, this structural effect is found overall when the number of matches and mismatches with the *model* is held constant across the target structures, and also when there is a total correspondence between the *model* and the initial syllable of the target bearing item. Contrary to the suggestion offered by Cutler Butterfield & Williams (1987), this result shows that the difference in RT introduced by the syllabic structure of the target bearing item cannot be attributed exclusively to the degree of correspondence between the *model* and the target item. Of course, we do not claim that *models* do not play a role. In fact, we found a small but significant effect of correspondence between the *model* and the target bearing item, although it is mostly present in items with a CV initial syllable.

Our results, then, are only in partial agreement with Cutler, Butterfield & Williams (1987). Like them, we found an effect of *model*, but unlike them, we found that it cannot account for the entire pattern of data. How could we account for the discrepant results between our experiment and that of Cutler *et al.*? One difference may be that in their study, the target was embedded in sentences, whereas we used lists of words. The different task demands and memory load of these two situations may be responsible for the greater impact of the *model* in the case of Cutler, Butterfield & Williams. Another reason may be the difference between English and French. Cutler *et al.*(1983, 1986, 1989) have suggested that the units used to perceive and process speech might vary from one language to another. In particular, they suggested that French subjects use a processing strategy based on syllables whereas English subjects do not. In this view, the impact of syllable structure in our first two experiments would be specific to French listeners.

In brief, the results obtained in the two preceding experiments indicate that, **at least for French**, the effect of syllabic complexity on phoneme monitoring time is robust and well- established. This effect cannot be reduced to the degree of correspondence between the *model* given to the subject and the structural organization of the stimulus item. Could it be, though, that the obtained complexity effects are representational in nature rather than due to processing? One thing to note is that in Experiment 2 there are a great many mismatches between the structure of the target and the *model*. That is, in only one case out of three, does the *model* match the actual syllabic structure of the target bearing item. If subjects come to notice this fact, they might choose to abandon the *model* they are given and revert to a default CV *model*, which is simpler, and which will not increase the average number of mismatches. Notice that the proportion of matches vs mismatches between *model* and target was the same in our study as it was in Cutler *et al.*'s study. So one cannot appeal to proportion of mismatches alone to account for the weak effect of *model* in French.

But there is one other reason to doubt that a *model* story, in and of itself, can account for the complexity effects. Even if complexity effects were due to a mismatch between a default CV representation and the stimulus, one would still be left with the following puzzle: detecting phonemes in disyllabic words starting with CV takes less time than in words starting with CVC (Experiment 2). This effect could easily be explained for monosyllables, where the end of the syllable is clearly marked by silence. But this is not so in disyllabic items, and both items (palace and palmier) start with the same sequence of three phonemes. For a *model* explanation to fly, one has to presuppose that the boundary of the first syllable is recovered from the signal prior to the matching procedure. Otherwise, both words would match a CV (or a CVC) mental *model* equally well. In that case, both words should yield the same reaction times. The fact then that the open/closed distinction has an impact on phoneme detection time even in disyllabic items means that the processing system provides a parsing into syllables to the post-perceptual mechanism responsible for the decision. But a syllabic parsing mechanism was precisely what the *model*-type explanation was trying to dispense with.

The upshot of this is that a *model* story cannot, on its own, account for the present set of data. One has to supplement the processing system with at least some notions about where syllable boundaries are likely to occur. However, the *model* hypothesis was just one of the two possible explanations for the existence of complexity effects. If we revert to a syllabic bottleneck hypothesis, everything works fine. This hypothesis states that, in order to detect an initial segment, the perceptual system first accumulates information that spans an entire syllable. One way to implement this hypothesis in a more concrete way would be to posit that the processing system categorizes speech inputs directly in terms of syllables, for instance by means of syllabic detectors. These detectors would gather information from the continuous acoustic input and generate a discrete code by selecting the best matching syllable with a mechanism that could be similar to that of lex-

Table 4

Duration (in ms) of the items used in Experiment 1 as a function of syllable structure. The obtained initial phoneme detection times are listed for comparison.

	Duration of items	Reaction times
CV	276	443
CCV	358	492
CVC	448	472
CCVC	540	519

ical access. Since phonemes are recovered after the syllable has been identified one can imagine many ways in which the properties of the syllable as a whole (such as syllable complexity, or syllable duration) can influence the detection time of its constitutent segments.

The strongest possible version of this model would state that no information can be recovered and used by higher levels until a unique syllable has been identified. However, a simple inspection of the reaction times obtained in this study suggests that this cannot be so: For instance, in Experiment 1, RTs are around 500 ms, yet the duration of monosyllables in isolation varies between 300 ms and 500 ms (see Table 4). Given a conservative estimate of 100 ms for motor response time, this means that on some occasions, subjects make their decision BEFORE the end of the syllable. More importantly, it is possible that they could have made their decision before they had enough information regarding the nature of the final consonant (Warren & Marslen-Wilson, 1987; 1988). This fact in itself seems to contradict the strong claim made above. It also raises a paradox: if subjects are responding before the last consonant, why do we find an effect of the presence or absence of that consonant?

This paradox can be resolved in a simple way. Recall that in standard analyses of RTs, we just use averages; so it could very well be that only the slow RTs are sensitive to the open/closed rime whereas the fast RTs are not. To test this hypothesis, we performed a number of reanalyses of the preceding two experiments, by splitting subjects into different groups according to their average speed of response.

The Role of Response Speed

A Reanalysis of Experiment 1

Mean reaction times were computed for each subject. Subjects were then distributed over four groups of ten according to their speed. The mean reaction time of the four groups was, respectively, 346 ms, 433 ms, 546 ms, and 639 ms. An analysis of variance was run, and in addition to the within-subjects factors of Onset Complexity and Coda, we introduced a new between- subjects factor: *Speed* (with four levels).

In this analysis, all three main factors were significant ($Onset\ Complexity : F(1,36) = 38.53, p < .001; Coda : F(1,36) = 14.38 p < .001; Speed : F(3,36) = 121.30 p < .001$). There was a significant interaction between Speed and Coda ($F(3,36) = 4.42, p < .01$). No other interaction was significant. To better understand the nature of the Speed and Coda interaction, we evaluated the effect of the type of rime across each of the four groups. Going from fastest to slowest subjects, the advantage of open versus closed syllables was -7 ms ($F < 1$), 11 ms ($F < 1$), 61 ms ($F(1,9) = 9.54, p < .02$) and 73 ms ($F(1,9) = 13.83, p < .005$) respectively. Thus, it is only significant for the 20 slowest subjects.[3] Figure 1 shows the pattern of complexity effects in the 20 fastest versus the 20 slowest subjects.

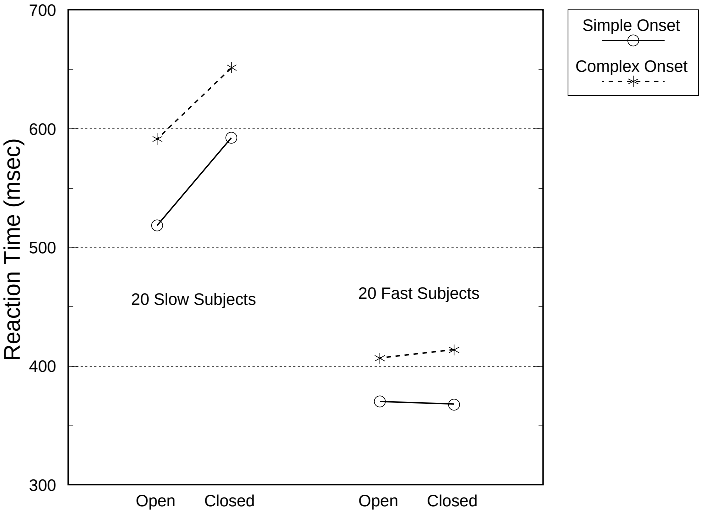

Figure 1. Reaction time to open and closed syllables with a simple or complex onset as a function of response speed (reanalysis of Experiment 1)

It could be argued that the disappearance of a Coda effect for fast subjects is only due to the fact that there is a general shrinking of the latencies when fast subjects approach the physiological limit (floor effect). However, this hypothesis can be rejected given the fact that (i) Onset Complexity does not interact with speed, and (ii), this effect was significant for all groups of subjects. Going from faster to slowest subjects, the Onset Complexity effect was 33 ms ($F(1,9) = 8.31, p < .02$), 50 ms ($F(1,9) = 8.71, p < .02$), 44 ms ($F(1,9) = 6.91, p < .03$) and 89 ms ($F(1,9) = 15.46, p < .004$) respectively. There was an increase in the magnitude of the Onset Complexity effect, but the effect is significant in the four groups of subjects.

---

[3] This phenomenon obtains for the structures with or without initial cluster: for the syllables with initial cluster (CCV and CCVC), the Coda effect is 7 ms (*F* 1) for the 20 fastest and 60 ms (*F* 1 19 7 81*p* 02) for the 20 slow subjects. For the syllables with cluster (CV and CVC), one goes from a Coda effect of -2 ms (*F* 1) for the fast subjects to an effect of 74 ms (*F* 1 19 13 20*p* 002) for the slow subjects. The interaction between the Speed factor (here with two levels) and Coda is significant in the case of syllables without initial cluster (*F* 1 38 10 19 *p* 003) and only marginal in the other case (*F* 1 38 3 65 05 *p* 1).

Discussion

The reanalysis reveals that the presence of an initial cluster does affect the performance of both fast and slow subjects. In contrast, the presence of a final consonant has a significant effect only for the slow subjects. This indicates that the two structural factors presented in Experiment 1 do not have the same status and do not contribute in the same fashion to syllabic complexity. Although the onset seems to be processed in a mandatory fashion, the coda in the rime is optionally taken into account. For fast subjects, it looks as if they were responding on the basis of ‘ *truncated*’ syllables, that is, only the onset plus nucleus part of the stimuli. Given their mean reaction times, this makes perfect sense. The 20 fastest subjects were responding with a mean latency of 390 ms, which corresponds to the duration of the longest open syllables but is 100 ms earlier than the end of the closed syllables. In other words, at the point when fast subjects made their decision to respond, the portion of signal distinguishing open and closed syllables might not yet even have been presented. It is not very surprising then that no effect of open versus closed was found for these subjects.

This result may sound relatively trivial, but it flies in the face of a model in which syllables function as a discrete stage (or bottleneck). In such a model, it should not be possible at all for subjects to respond before the end of a syllable. By definition, no information regarding the syllable’s internal structure should be accessible until a unique syllable has been identified. Of course, one should realise that the stimuli were monosyllables spoken in isolation, and hence they had a much longer duration than they would have had in a more natural context. So it remains possible that fast subjects engage in an ad-hoc strategy for these unusually long stimuli, a strategy that would not be available if the syllables were shorter. This is why, before going into a discussion of the implication of this ‘*truncation*’ phenomenon, we would like to know if it also obtains for the disyllabic items used in Experiment 2. Given that the first syllables of such items are much shorter than when they are produced as monosyllables (about twice as short), one should expect much less impact of speed differences in these cases. The following is a reanalysis of Experiment 2.

A Reanalysis of Experiment 2

As will be recalled, Experiment 2 was designed in part to investigate the effect of the model given to subjects on phoneme detection time in three syllabic structures, CV, CVC and CCV. In the following, we will ignore the *model* aspect of this experiment by pooling together all the reaction times, irrespective of the nature of the model.

Subjects were again divided into groups of 10 according to their mean reaction times. This yielded three groups with mean reaction times of, respectively 277 ms, 405 ms and 488 ms respectively. An analysis of variance with a between-subject speed factor plus a within- subject Complexity factor was conducted. The two main factors were significant ($\mathit{Complexity}F(2,54) = 17.10, p < .001, \mathit{Speed}F(2,27) = 65.29, p < .001$). The interaction between the two factors was not significant ($F(4,54) = 2.01, p = .1$); however, if one restricts the analysis to the two extreme groups, the interaction between Complexity and Speed becomes significant ($F(2,36) = 4.59, p < .02$).

We computed a series of contrasts in the three groups. The contrast between CV and CCV was significant in the three groups (starting with the fastest: 61 $ms, F(1,9) = 25.69, p < .001;$ 36 $ms, F(1,9) = 5.35, p < .05;$ 37 $msF(1,9) = 6.35, p < .04$). However, the contrast between CV and CVC was not significant in the fastest subjects (6 $ms, F < 1$), nor in the medium subjects (13 $ms, F < 1$), but it was significant for the slowest subjects (32 $ms, F(1,9) = 5.86, p < .04$).

Discussion

Initial phoneme detection in disyllabic words is sensitive to the presence of an initial cluster. This effect is robust and is found consistently irrespective of the speed of response. In contrast, the presence or absence of a coda consonant only yields strong effects in the slowest subjects. Such a result obtains even with initial syllables of disyllabic items which are much shorter than monosyllabic items (around 200 ms). It is probable, then, that even the fastest subjects can prepare their response while the entire syllable is available. However, these subjects do not seem to use this information but persist in taking into account only about the first half of the syllable.[4]

However, proponents of a bottleneck account could still point out two shortcomings in the previous reanalyses.

First, our conclusions (about the absence of a syllabic bottleneck) were only reached in a post-hoc fashion. There was no *experimental* manipulation of subjects’ speed, but just a post-hoc selection of some subjects. Hence we do not know whether response speed was the critical parameter, or whether the selected subjects would simply never show coda effects, even if their responses were slow. This is why we wanted to replicate the above findings by *experimentally* speeding up subjects selected at random.

A second point relates to the reasoning we used to assess that responses were triggered before a whole syllable had been processed. It was based on two arguments. The first argument relies on the comparison of mean response times with the physical duration of the stimuli (as in Table 4). When response times were shorter than the duration of the syllable, it was argued that subjects responded before having identified the syllable, hence disproving the syllabic bottleneck hypothesis. Although such considerations might have some heuristic value, a knock down argument is difficult to make. In fact, it might be that the *identification point* of some syllables occurs **before** stimulus offset, making it possible that syllables were identified before phoneme response were made. An argument solely based on stimulus durations is difficult to make in the absence of an independant way of measuring information flow at the various levels at hand. The second argument was based on the presence or absence of a coda effect on

---

[4] It remains interesting to know precisely how much is taken into account: it is the first two segments, everything up to the nucleus? More research is needed to draw more precise conclusions.

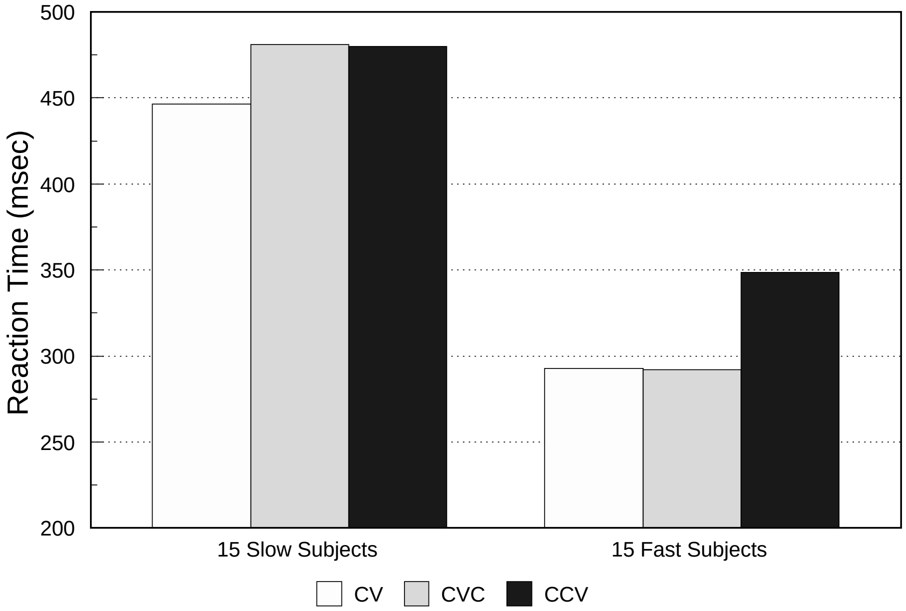

Figure 2. Reaction time to disyllabic items starting with CV, CVC, or CCV as a function of response speed (reanalysis of Experiment 2)

phoneme detection times. When a coda effect was observed, the inference was that the whole syllable had been taken into account. Conversely, when no coda effect emerged, we concluded that only the initial part of the syllable had been processed. However, this latter inference might not hold. It might be that fast reponses are typical of a response strategy whereby no coda effect obtains at all, even though the reponse might be based on the entire syllable (e.g. fast responders can better focus their attention at the onset of syllabic structures and ignore the structure of the rime). To address these potential shortcomings, we need to use another way to assess how much of the stimulus has been processed by the time a response is given.

This is why we ran an additionnal experiment where we asked subjects to detect phonemes as fast as they could. In order to test whether or not the entire syllable has been processed when the response is being made, we measured whether the lexical status of the syllable would matter.

Experiment 3

The aim of this experiment is to investigate the reliability of the interaction between speed of response and complexity effects, namely, the fact that the processing system can extract individual phonemes before an entire syllable has been analysed. In the following, we exploit the fact that in monosyllabic items, phoneme detection times tend to be influenced by lexical variables. For instance, frequent items show faster RTs than infrequent ones (Dupoux & Mehler, 1990; Eimas, Marcovitz Hornstein & Payton, 1990). Monosyllabic words yield faster RTs than otherwise similar non-words (Rubin, Turvey, & Van Gelder, 1976; Cutler *et al.*, 1987; Eimas *et al.*, 1990). The prediction is, therefore, quite straightforward. If in the case of fast responses, CVC syllables go through a stage where only the CV portion is available, then this CV portion should not be sufficient to complete lexical access (provided that the identification point comes after the vowel). Hence, when subjects are asked to respond very fast, no lexical effects should be found with CVC items, whereas those effects should still be found with open syllables. This would be similar to the disappearance of lexical effects on phoneme identification with fast reponses (Fox, 1984; Miller & Dexter 1988).

To test this, we took the materials used in Experiment 1 of Cutler *et al.*(1987). In this experiment, three structures were tested and crossed with the lexical status of the item: CV, CVC and CCV. In the original experiment, the authors found a lexicality effect for the three structures. In this replication, instructions to subjects stressed that they should try to respond as fast as they could, if possible before the end of the items. The prediction was that, if subjects can ignore the final part of the stimuli provided that the response is fast enough, there should be a lexical effect for CV and CCV, but not for CVC. Moreover, there should be an onset complexity effect but not a coda effect.

Method

Materials and procedure.

The same materials as in Cutler *et al.*(1987) Experiment 1 were used. They consisted of 25 pairs of matched words and non-words. Five pairs had CV structure (mean frequency 27200 per million), 10 pairs were CVC (mean freq. 1830) and

10 pairs were CCV (mean freq. 1940). In the CV and CCV pairs, the non-words had the same segments as the words, except for the first phoneme (e.g: /*pi*/-/ *ti*/, / *gl*ɑ/-/ *kl*ɑ/). In the CVC pairs, it was the final consonant that carried the change (e.g. / *d*ɑ*t*/-/ *d*ɑ*k*/). For the CCV items, the first consonants used in the words and non-words were globally counterbalanced so that a potential lexical effect could not be attributed to low-level factors (that is, matched to the word-nonword pair */gl*ɑ*/-/kl*ɑ*/*, there was the pair */krø/-/grø/*). For CV items, this counterbalancing was only done in terms of manner of articulation of the first consonant. There were 324 additionnal filler items (60% words, 40% non-words, 75% disyllabic, and 25% monosyllabic).

The 374 items were distributed into 100 sequences ranging in length from 1 to 6 items. A specific target phoneme was associated to each of these sequences. Ten sequences, each 6 items in length, did not contain an occurrence of the specified target. Forty sequences had targets occurring in first, second, or sixth position. The remaining 50 sequences each had one of the 50 experimental items occurring in third, fourth, or fifth position. Each word was matched with its paired non-word on position in the list. Each list ended with a short beep. The lists were put into two separate blocks of 50, that would be presented either in the AB or in the BA order. To keep the present experiment as similar as possible to the original Cutler *et al.*'s experiment, the target phonemes were presented visually. Before each list, the target phoneme was presented on a computer screen with three ‘models’ (town names) of a different syllabic structure (e.g. “/p/ as in Paris, Prague, Perpignan”). The materials were split into two main blocs and the order in which they were run was counterbalanced across subjects. The auditory materials were digitized from the original tapes that were used in Cutler *et al.*(1987) Experiment 1. Each of the short lists was digitally cut into a separate file, and reassembled in the same order as in the original experiment. The words were presented at the approximate rate of one word every 1 second and a half, and a pause of 3 seconds was inserted at the end of each list.

Subjects. Thirty students from the Faculty of Psychology, University of Paris V were tested . They were assigned to one either the AB or the BA group. The Subjects' task was to press a button as fast as they could as soon as they heard the previously specified phoneme at the onset of an item. Instructions strongly emphasized the necessity of fast responding.

Results

Reaction times above 1000 ms and under 100 ms were removed. This was less than 1% of the data. Globally, the error rate was 2.58%. Mean reaction time was 365 ms, which was about 100 ms faster than RTs obtained by Cutler *et al.*(1987) with the same materials. Reaction times for the two groups of subjects (AB and BA) did not differ significantly and did not interact with the other factors; hence in the following we only considered a single homogeneous group of subjects. The mean reaction times for each condition are shown in Table 5.

Table 5
Phoneme detection times for CV, CVC and CCV words and non-words (experiment 3).

Syllabic structure	Non-word	Word	Lexical Effect
CV	370 (64)	332 (49)	38
CVC	341 (68)	333 (65)	8
CCV	387 (71)	367 (60)	20

Two analyses of variance were conducted, one with subjects and another with items as random variables. The factor of lexical status yielded a significant effect in both analyses: words gave rise to faster RTs than non-words ($19\ ms; F_1(1,29) = 14.57 p < .001; F_2(1,22) = 4.71, p < .05; min\ F\ (1,36) = 3.56, p = .07$). The syllable structure factor only had a significant effect in the subjects analysis only ($F_1(2,58) = 18.31, p < .001; F_2(2,22) = 3.22, p = .06$). The interaction between these two factors was nearly significant in the subject analysis ($F_1(2,58) = 3.08, p = .053; F_2 < 1$). Such a trend suggests that the lexical status effect has a differential impact according to the syllabic structure of the target bearing item.

To test the hypothesis that lexical effects are not homogeneous across syllabic structure, we analyzed it separately for CV, CVC and CCV items. It turned out that the effect of lexical status was significant in the subject analysis for CV items ($38\ ms; F_1(1,29) = 10.99, p < .0025; F_2(1,4) = 1.95 ns$) and for CCV items ($20\ ms; F_1(1,29) = 4.98, p < .03; F_2(1,9) = 1.98 ns$). However, no significant effect of lexical status was found in CVC items ($8\ ms; F_1(1,29) = 1.2, ns; F_2 < 1$). When all the items with an open rime (CV & CCV) were pooled, the effect of lexical status was significant by subjects, and almost reached significance in the analysis by items ($26\ ms; F_1(1,29) = 15.1, p < .001; F_2(1,13) = 4.48, p = .054$).

The above analysis suggests that open and closed items behave differently with respect to lexical effects: closed items do not show a lexical effect whereas open items do. This ought to generate an interaction between lexical status and the Coda factor (i.e. whether the syllable was open or closed). Indeed, we found a significant interaction in the subject analysis ($F_1(1,29) = 6.06, p < .02$). However, since the item analysis did not reach significance, one has to be cautious in trying to generalize the above interactions to other materials.

Finally, we ran individual comparisons of the RTs to the different structures. CV items yielded faster RTs than CCVs ($26\ ms; F_1(1,29) = 13.32, p < .001; F_2(1,13) = 1.8, ns$), CVCs yielded faster RTs than CCVs ($40\ ms; F_1(1,29) = 46.87, p < .001; F_2(1,18) = 6.39, p < .05; minF\ (1,23) = 5.62, p < .03$). In contrast, the difference between CV and CVC items was small and only marginally significant ($-14\ ms; F_1(1,29) = 4.10, .1 < p < .05; F_2 < 1, ns$). One should note that the latter trend was that CVC were faster than CV, a pattern opposite of the classical complexity effect found by Cutler *et al.*(1987) using the same materials. Finally, items with a simple onset yielded significantly faster RTs than items with a complex onset ($F_1(1,29) = 37.63, p < .001; F_2(1,22) = 5.04, p < .04; minF\ (1,28) = 4.44, p <$

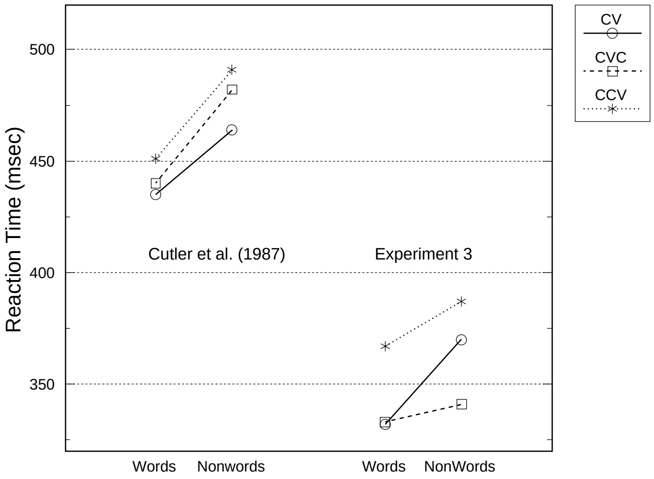

Figure 3. Reaction time to CV, CVC and CCV words and nonwords. The left panel shows the results obtained in Cutler *et al.*(1987), the right panel shows our replication (Experiment 3).

.05).

Discussion

In this experiment, we observed the following three results (see Figure 3):

First, the RTs we found were 100 ms faster than the RTs found in the original experiment by Cutler *et al.*(1987), even though the stimuli were physically identical. The only difference between our Experiment 3 and their Experiment 1 is that we put very strong emphasis on response speed.[5] We then showed that it is possible to obtain very short RTs (as short as the ones typically obtained in syllable detection experiments), while preserving a significant lexical status effect.

Second, a complexity effect was observed. However, this was mainly due to CCV items producing longer RTs than CV or CVC. Such a robust effect of initial cluster is consistent with our Experiments 1 and 2. The presence or absence of a coda consonant, however, did not have the effect that was found in the original study (Cutler *et al.*, 1987). CVC items thus yield faster reaction times than they should. Again, it is as if CVC targets were processed like CVs.

Third, the effect of lexical status was only significant for the open syllables CV and CCV. CVC items did not show any trace of a lexical effect, which gave rise to a significant interaction between presence/absence of a coda and lexical status. Such a result differs from the original Cutler *et al.* study, where a lexical status effect was found for the three structures (CV: 29 ms, CVC: 42 ms, CCV: 40 ms).

How can we account for these divergent data patterns? Here is our hypothesis: in the present experiment, subjects were told to respond *extremely fast.* Their average RT being around 355 ms, they could only base their decision on roughly the first 255 ms of the signal. This period mostly comprises the onset and the nucleus of the syllable. In these truncated syllables, there is still a structural difference between CV and CCV syllables. Moreover, the information about the onset and the nucleus is sufficient to initiate lexical access. So CV and CCV words will have a processing advantage over CV and CCV non-words. In contrast, in the case of a truncated CVC, the initial part (CV) was identical for a word and its matched non-word. In that case, even if the truncated part attempted to trigger lexical access, that attempt would have yielded the same result for words and non-words. Hence there was no effect of lexical status for these items.

These results are compatible with our previous reanalyses of Experiments 1 and 2. They give converging evidence showing that by speeding up responses, one can induce subjects to respond on the basis of fragmentary acoustic information belonging to the initial part of the syllable.

General discussion

Initial phoneme detection time depends on the complexity of the syllable in which it appears. In Experiments 1 and 2, we showed that both the complexity of the onset and the presence of a coda contribute to syllabic complexity and affect phoneme detection latencies. This was borne out by the following ranking of reaction times: CV $<$ CVC $<$ CCV $<$ CCVC. In Experiment 2, we argued that this complexity effect is also found for the first syllable of polysyllabic items, which suggests that it cannot solely be due to differences in stimulus duration. Also, we found that the complexity effect cannot be solely explained with reference to the representation used by subjects to hold the phoneme target in memory (the model), but that at least part of the complexity effect has to be due to some early process. The interpretation that was proposed is that perceptual processes underlying the extraction of individual segments have to rely on information distributed over a comparatively large chunk of the signal, possibly over an entire syllable. Viewed in this way, the complexity effect would be very similar to other findings suggesting that the quality and length of the V in a CV stimulus are used to recover the identity of the C (Diehl *et al.*, 1987; Miller, 1981). What our results add to this picture is that the processing system needs to take into account information on adjacent segments **which belong to the same syllable**. This is illustrated in Experiment 2, where, for instance both *balance* and *balcon* start with the same three segments /bɑl/. Yet, the /l/ in *balcon* belongs to the first syllable whereas the /l/ in *balance* belongs to the second syllable. This difference in structure is reflected in the data by the fact that *ba.lance* gave rise to faster RTs than *bal.con*. The bottomline is that to account for these data, one might have to posit that the syllable structure is recovered at a comparatively early stage in perception.

However, subsequent reanalyses of Experiments 1 and 2, plus the results of Experiment 3, indicate that such an interpretation has to be further elaborated. In the reanalyses of Experiments 1 and 2 we found that whereas the effect of onset seems to be very robust and is found uniformly across all subjects, the effect of codas only shows up in slow subjects. This latter finding is consistent with studies showing that with the

---

[5] Also, due to the computerized presentation of the target phonemes, the experiment was 20 minutes shorter than the original one.

fragment detection technique (Mehler *et al.*, 1981), syllabic effects disappear when subjects' responses are very fast (Sebastian *et al.*, 1992). This suggests that syllables are not processing bottlenecks and that subjects can respond before having identified the full syllable.

In further support of this, we can quote a study by Miller & Dexter (1988) who found that response times interacted with the effect of vowel duration on initial consonant phonetic boundaries. When subjects make non-speeded classification responses, vowel duration reliably shifts the phonetic boundary in CV stimuli as in the original study of Miller & Liberman (1979). However, when subjects respond in a speeded fashion, the vowel duration effect disappears; moreover, the phonetic boundary seems to correspond to that of the shorter vowel duration, as if the perceptual system were delivering a percept based on the vowel duration known at the time the subject decides to respond. A similar phenomenon might be present in the study of Gordon *et al.*(1993) who found that the cues that influenced the classification of a consonant continuum changed when a secondary distractor task was added. Although the authors accounted for their data in terms of attentional influence on phonetic perception, it is very likely that the secondary task had a major effect on response time (since a lot of emphasis was put on accuracy as opposed to speed). Interestingly, it is with such slowed down responses that subjects showed the greater influence of later occurring cues (such as formant transitions for consonants and vowel duration for vowels) as opposed to earlier occurring cues (burst and formant composition, respectively). It remains to be seen whether such effects are really due to speed differences rather than, as the authors claim, to attentional differences. In brief, our results are compatible with a growing body of evidence suggesting that even though phonetic cues are integrated over a large chunk of signal, there is no processing bottleneck. The current evaluation of a segment's identity is available and usable at any given point. In other words, the world really could be what it looks like from the viewpoint of the gating paradigm. Yet this intepretation is not as straightfordward as it may seem. In the following, we discuss three possible interpretations of these results in terms of the information flow over large-sized units.

The first interpretation would question the generality of the response speed effects. It could be that such an effect is an artifact of the very slow rate of presentation of the isolated stimuli used in psycholinguistic experiments. Suppose that one artificially expands a stimulus in order to make it, say, 10 seconds long. No one would expect subjects to wait for the end of such a stimulus before responding. So it could be that in fact there are processing bottlenecks in running speech, but that listening to isolated lists of words recorded in a soundproof room enables subjects to by-pass the syllabic level altogether and to rely on acoustic cues to perform the task at hand. The way to test this would be to examine the effect of syllabic complexity with connected speech spoken at a natural rate. Under the above "ecological" hypothesis, subjects should always respond on the basis of entire syllables, that is, "truncated" responses, if any, should yield chance performance. If, on the other hand, "truncation" is a general phenomenon in speech perception, we should expect the same interaction between syllable complexity and response speed that we documented in Experiments 1, 2 & 3, even with fast and heavily coarticulated speech.

The second interpretation holds that the interaction of response speed on syllabic effects is a general phenomenon that reveals important aspects of normal speech perception. The prediction here, is that, even with noisy input and running speech, one could still find conditions where subjects could respond before a full syllable has been extracted. If this is so, one would then have to say that even though syllables do constitute a processing level in speech comprehension, they do **not** constitute a processing bottleneck. The idea would be that, for instance, syllabic candidates are continuously evaluated as time passes and information accumulates. Such an evaluation will eventually stabilize once the entire syllable has been presented. An illustration of such an idea can be given in cascade processing models such as McClelland & Elman, (1986) TRACE model (although only the bottom-up parts of the model is relevant to our discussion). One could propose a level of syllabic units somewhere between the phonetic level and the lexical level, but because of the dynamics of the network, the information flow would still be gradual and incremental. Of course, we have not demonstrated that the information flow is completely continuous. In fact, we have found that even for very fast responses there is still an effect of onset complexity. So it could be that there is a processing bottleneck, only one spanning over smaller units. In order to illustrate how such a revised proposal could be implemented, one can imagine that acoustic/phonetic encoding proceeds with a unit the size of half syllables (that is the transitions from initial consonant(s) to nucleus or vice versa; see Fujimura, 1976; Samuel, 1989 for a similar proposal). All segments inside such units would be hypothetically perceived in an 'integral' fashion. Each of these half syllables, activated in sequence, would be used to construct a structured phonological representation on the basis of which phoneme or fragment detection tasks are performed (Dupoux & Mehler, 1992). So when subjects respond very rapidly, only half a syllable is processed, and responses are sensitive to the acoustic/phonetic encoding within this unit. When subjects respond slowly, a complete syllabic structure has been processed, which results in syllable-wide complexity or frequency effects. Further research should help us untangle these issues.

Between these two extremes, we have a third intermediate possibility. It may be that the information flow is cascadic in some part of the system, but discontinuous in others. In particular, one could say that information is allowed to cascade up to the decision system that allows subjects to perform detection tasks. However, the same information would be blocked, say, for the purpose of lexical access. Indeed, our theoretical argument about the interest of having a large-sized bottleneck for the purpose of word recognition still holds. In other words, what is wrong with 'truncation' is not that the stimuli are unnatural, but that the task is unnatural. One way to test this is to look for an effect of the lexical status of the truncated part of a CVC item. So, for instance, both /pit/ and

/pɛt/ are words, but /pi/ is a word whereas /pɛ/ is not. The prediction of a complete cascade model is that there should be an RT difference between these two stimuli, but only for fast responses (see Christophe (1993) for some indication that such cascade effects do not occur).

In brief, the phoneme detection technique has proved to be sensitive to large sized units such as syllables. Moreover, the effect of syllabic structure provides an interesting puzzle. On the one hand, certain aspects of the data suggest that syllables have a perceptual locus, and can help to integrate many psychoacoustical findings showing context dependence in phonetic perception. On the other hand, although syllable effects are robust, they are not mandatory in the sense that subjects can respond on the basis of syllable fragments, provided that the response is fast enough. We have outlined three ways in which such a time course of structural effects could be accounted for. All of these alternatives need considerable refinement and empirical support, but it should be noted that all of them require taking into account syllable-like units in some way. Models that only incorporate a segmental representation have to be supplemented with larger units like syllables. What remains unclear is the level at which these units play a role (e.g. stimulus encoding versus representation in memory or both), and how they relate to the unfolding of speech information through time. We hope to have demonstrated that phoneme monitoring techniques are useful to formulate hypotheses about such questions and can help to uncover the architecture of prelexical processing.

References

Bailey, P.J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of /s/-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 536-563.

Christophe, A. (1993). *Rôle de la prosodie dans la segmentation en mots*. Unpublished Doctoral Dissertation, EHESS, Paris.

Cutler, A., Butterfield, S., & Williams, J.N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language*, **26**, 406-418.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language specific comprehension strategy. *Nature*, **304**, 159-160.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385-400.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, **19**, 141-177.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1989). Limits on bilinguism. *Nature*, **320**, 229-230.

Diehl, R.L., Kluender, K.R., Foss, D.J., Parker, E.M., & Gernsbacher, M.A. (1987). Vowels as islands of reliability. *Journal of Memory and Language*, **26**, 564-573.

Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, **29**, 316-335.

Dupoux, E. & Mehler, J. (1992). Unifying awareness and on line studies of speech: a tentative framework. In J. Alegria, D. Holender, J. Morais, & M. Radeau (Eds.) *Analytic approaches to human cognition*. pp 59-76. Elsevier, The Netherlands.

Dupoux, E. (1993). Prelexical processing: the syllabic hypothesis revisited. G. Altmann & R. Shillcock (Eds.) *Cognitive Models of Speech Processing.*. pp 81-114. Erlbaum, U.K.

Eimas, P.D., & Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors, *Cognitive Psychology*, **4**, 99–109.

Eimas, P.D., Marcovitz Hornstein, S.B. & Payton, P. (1990). Attention and the role of Dual Codes in Phoneme Monitoring. *Journal of Memory and Language*, **29**, 160-180.

Foss, D.J., & Gernsbacher, M.A. (1983). Cracking the dual code: Toward a unitary model of phoneme identification. *Journal of Verbal Learning & Verbal Behavior*, **22**, 609-632.

Fowler, C.A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.

Fox R.A. (1984). Effects of lexical status on phonetic categorisation, *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 526-540.

Fujimura, O. (1976). Syllables as concatenated demisyllables and affixes. *Journal of the Acoustical Society of America*, **55**.

Gordon, P.C., Eberhardt, J.L., Rueckl, J.G. (1993). Attentional modulation of the phonetic significance of acoustic cues, **25**, 1-42.

Klatt, D. H. (1980). Speech perception: A model of acoustic- phonetic analysis and lexical access, In R. A. Cole (Ed), *Perception and production of fluent speech*, Erlbaum: Hillsdale, N.J..

Marslen-Wilson, W.D. (1984). Functions and process in spoken word recognition. In H. Bouma, D.G. Bouwhuis (Eds.), *Attention & Performance X: Control of language process*, Hillsdale, N.J.: Erlbaum.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of Speech Perception. *Cognitive Psychology*, **18**, 1-86.

Mehler, J. (1981). The role of syllables in speech processing: Infant & adult data. *Philosophical Transactions of the Royal Society*, **B 295**, 333-352.

Mehler, J., Dommergues, J.Y, Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation, *Journal of Verbal Learning and Verbal Behavior*, **20**, 298-305 .

Miller, J.L., & Dexter, E.R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 369-378.

Miller, J.L., & Liberman, A.M. (1979). Some effects of later-occurring information on the perception of stop consonants and semivowels. *Perception & Psychophysics*, **25**, 457-465.

Miller, J.L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas & J.L. Miller (Eds.), *Perspectives on the study of speech*, Hillsdale, N.J.: Erlbaum.

Miller, J.L. (1986). Limits on later occurring rate information for phonetic perception. *Language and Speech*, **29**, 13-25.

Mills, C.B. (1980). Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 528-535.

Pisoni, D.B., & Luce, P.A. (1987). Acoustic-Phonetic representations in word recognition. *Cognition*, **25**, 21-52.

Repp, B.H., Liberman, A.M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricative manner. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 621-637.

Rubin, P., Turvey, M.T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics*, **19**, 394-398.

Savin, H., & Bever, T. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, **9**, 295-302.

Samuel, A.G. (1989) Insights from a failure of selective adaptation : Syllable-initial and syllable-final consonants are different, *Perception and Psychophysics*, **45**, 485-493 .

Segui, J., Dupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation and lexical access. In J. Altmann (Ed.), *Cognitive Models of Speech Processing*, Cambridge, Mass: MIT Press.

Segui, J., Frauenfelder, U., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring & lexical access, *British Journal of Psychology*, **72**, 471-477.

Swinney, D.A., & Prather, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, **27**, 104-110.

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

Tomiak, G.R., Mullenix, J.W., & Sawusch, J.R. (1987). Integral perception of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, **81**, 3,

Treiman, R., Salasoo, A., Slowiaczek, L.M., & Pisoni, D.B. (1982). Effects of syllable structure on adults' phoneme monitoring performance. *Progress Report nb 8*, Indiana University, Speech Research Laboratory.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, **41**, 3, 262-275.

Warren, P., & Marslen-Wilson, W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics*, **43**, 21-30.

Wood, C.C., & Day, R.S. (1975). Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception and Psychophysics*, **17**, 346-350.

Materials

Materials used in Experiment 1.

CV	CCV	CVC	CCVC
Voiced initial phoneme			
bi	bri	bid	brit
bo	blo	bog	brok
da	dra	dag	drab
de	dre	deb	drep
Unvoiced initial phoneme			
pa	pla	pad	prak
po	pro	pob	prot
ti	tri	tig	trid
tu	tru	tuk	trup

Materials used in Experiment 2.

CV	CVC	CCV
palace	palmier	placard
pourris	pourboire	prouver
garage	gardien	gravier
touriste	tournée	troupeau
barreau	barbu	bravoure
burin	burlesque	brutal

Materials used in Experiment 3

CV Items					
Non-words	da	ti	na	gu	co
	/dɑ/	/ti/	/nɑ/	/gy/	/ko/
Words	tas	pis	ma	du	do
	/tɑ/	/pi/	/mɑ/	/dy/	/do/
CVC Items					
Non-words	poc	pit	tal	berre	tote
	/pɔk/	/pit/	/tɑl/	/bɛr/	/tɔt/
	buk	cak	dac	gasse	posse
	/byk/	/kɑk/	/dɑk/	/gɑs/	/pɔs/
Words	pote	pic	tard	belle	toc
	/pɔt/	/pik/	/tɑr/	/bɛl/	/tɔk/
	but	cap	date	gaffe	poche
	/byt/	/kɑp/	/dɑt/	/gɑf/	/pɔʃ/
CCV Items					
Non-words	pru	bré	tra	dro	tru
	/pry/	/bre/	/trɑ/	/dro/	/try/
	dri	cla	glo	cra	greu
	/dri/	/klɑ/	/glo/	/krɑ/	/grø/
Words	bru	pré	drap	trop	dru
	/bry/	/pre/	/drɑ/	/tro/	/dry/
	tri	glas	clos	gras	creux
	/tri/	/glɑ/	/klo/	/grɑ/	/krø/