Grapheme context effects on phonemic processing

DIJKSTRA, Ton, SCHREUDER, Robert, FRAUENFELDER, Ulrich Hans

Abstract

This study investigated the structural and temporal relationship between graphemic and phonemic processing by means of a cross-modal priming procedure. Subjects made a forced choice on the identity of the vowel in an auditorily presented syllable. To determine if and when phonemic representations are automatically activated by graphemes, visual letter primes were presented before, during, or after presentation of the syllable. In the indirect priming condition, the relationship (congruent/incongruent) between the letter and the consonant adjacent to the vowel was manipulated, whereas in the direct priming condition that between the letter and the target vowel itself was varied. In two indirect priming experiments faster reaction times were obtained over the entire range of SOAs tested when the prime was congruent with the consonant of the syllable, than when it was not. In a third direct priming experiment SOA-dependent facilitation effects were found with respect to a bimodal baseline-condition when the prime was congruent with the target vowel, and inhibition effects when it was congruent with the competing target […]

Reference


DOI: 10.1177/002383098903200201

Available at:
http://archive-ouverte.unige.ch/unige:83836

Disclaimer: layout of this document may differ from the published version.
GRAPHEME CONTEXT EFFECTS ON PHONEMIC PROCESSING*

TON DIJKSTRA  
NICI, University of Nijmegen  
and  
Max-Planck-Institut für Psycholinguistik, Nijmegen  
ROBERT SCHREUDER  
Interfaculty Research Unit for Language and Speech  
University of Nijmegen  
and  
ULI H. FRAUENFELDER  
Max-Planck-Institut für Psycholinguistik, Nijmegen

This study investigated the structural and temporal relationship between graphemic and phonemic processing by means of a cross-modal priming procedure. Subjects made a forced choice on the identity of the vowel in an auditorily presented syllable. To determine if and when phonemic representations are automatically activated by graphemes, visual letter primes were presented before, during, or after presentation of the syllable. In the indirect priming condition, the relationship (congruent/incongruent) between the letter and the consonant adjacent to the vowel was manipulated, whereas in the direct priming condition that between the letter and the target vowel itself was varied. In two indirect priming experiments faster reaction times were obtained over the entire range of SOAs tested when the prime was congruent with the consonant of the syllable, than when it was not. In a third direct priming experiment SOA-dependent facilitation effects were found with respect to a bimodal baseline-condition when the prime was congruent with the target vowel, and inhibition effects when it was congruent with the competing target vowel. The results support the hypothesis of automatic grapheme-to-phoneme activation before word recognition.

Key words: cross-modal, grapheme-phoneme activation, sublexical, priming

INTRODUCTION

The processes underlying listening and reading necessarily differ in their early stages, due to differences in sensory organs and input signals. Nevertheless, since the same

* This research is part of a thesis project of the first author supported by the Max-Planck-Institut für Psycholinguistik and the University of Nijmegen. The authors wish to thank Agnes Bolwiender and Hans Oskamp for their assistance in running and analyzing the experiments. They also wish to thank Vicki Hanson, Bruno Repp, Dirk Vorberg, and Pienien Zwitserlood for comments on earlier versions of this paper.

Address all correspondence to: Ton Dijkstra, NICI, P.O. Box 9104, NL-6500 HE Nijmegen, The Netherlands; or to email-address dijkstra@hnykun53.bitnet.
message can be understood in both modalities, the two processes must also converge at some point. In fact, most researchers would agree that reading and listening share the same database, the mental lexicon. This paper examines a more controversial possibility, that of contact between visual and auditory representations at a lower level of representation. The smallest representational units for which such contact would seem feasible, are those of letters (graphemes) and speech sounds (phonemes).

Grapheme—phoneme contacts play an important role in several models of visual word recognition, such as the dual-route model (e.g., Coltheart, 1978, 1980). The dual-route model assumes that word recognition may proceed via either a direct or an indirect processing route. The first operates by mapping a word's extracted visual features directly onto its stored lexical representation; the second by translating the word's orthographic code into a phonological code and by subsequently using this phonological code to access the lexicon. It has been suggested (Coltheart, 1978; Venezky, 1970) that readers make use of an internal set of grapheme—phoneme correspondence rules (GPCs) that converts single letters or letter clusters into the corresponding phonemes or phoneme clusters. To account for the finding that effects of phonological recoding are found only under specific conditions (e.g., when pronouncing low-frequency words (Seidenberg, 1985a)), the dual-route model assumes that the direct route usually wins the race to word recognition.

The assumption of a strong version of a dual-route theory that the indirect route alone suffices for word recognition, has been heavily criticized (e.g., Humphreys and Evett, 1985). Indeed this route runs into serious difficulties because there are many exceptions to the GPC-rules. Many correspondences between spoken and written words are inconsistent (e.g., the pronunciation of -AVE in HAVE) and even arbitrary (e.g., -OLO in COLONEL). This is due in part to the fact that writing systems reflect not only phonological principles, but also morphological and etymological ones.

The time-course model of visual word recognition (Seidenberg, 1985b, 1985c, 1987; Seidenberg, Waters, Barnes, and Tanenhaus, 1984) circumvents this problem by abandoning the assumption that the two routes function independently. In an implemented version of this model (Seidenberg and McClelland, 1987) orthographic representations are not “translated” or “recoded” into a phonological format, but activation is spread from orthographic to associated phonological representations. The additional time required to build up activation of phonological as compared to orthographic representations is taken to explain the relatively small influence that the phonological characteristics of a word have on visual word recognition. The specific patterns of connections between letter and sound strings are established via an associative learning process. This process leads to general rule-like behavior, while allowing irregularly spelled words to be learned and recognized.

The theoretical assumption that phonological information is activated during visual word recognition is empirically corroborated by experiments that rely upon the visual presentation of word and nonword stimuli. A number of lexical decision studies (Rubinstein, Lewis, and Rubinstein, 1971; Coltheart, Davelaar, Jonasson, and Besner, 1977), for example, have shown reaction time (RT) differences between nonwords that are homophonous with real words (“pseudohomophones”: e.g., BRANE) and non-homo-
phonic nonwords (BRAME). It is argued that pseudohomophones take longer to reject as nonwords because the homophonic words become activated via their nonlexically assembled phonological code.

Humphreys and Evett (1985) signal two problems with this conclusion. First, the effects may be due to orthographic rather than to phonological similarity of nonwords to words, since most studies have not satisfactorily partialed out these two variables. Second, it has not been established that phonological representations of nonwords are assembled without the use of lexical knowledge (e.g., Marcel, 1980).

Some recent studies have investigated the issue of phonological activation in visual word recognition with tasks other than lexical decision. Van Orden and his colleagues (Van Orden, 1987; Van Orden, Johnston, and Hale, 1988) had subjects decide whether a visually presented word belonged to a prespecified semantic category. Errors occurred more frequently with homophones of category instances (e.g., ROWS, homophonic to ROSE of the category "A FLOWER") than to spelling controls (ROBS). An elevated error score virtually identical to that of homophones was found for matched nonword homophone foils (such as ROAS). These findings demonstrate that the phonological properties of words mediate their recognition in a categorization task that requires reading for meaning. They further suggest that the phonological activation occurred automatically, since such activation was actually detrimental to performance in this situation.

A different approach was taken by Perfetti, Bell, and Delaney (1988), who asked subjects to identify briefly presented lower-case target words that were followed first by an upper-case pseudoword mask and subsequently by a pattern mask (a row of Xs). They varied the orthographic and phonological properties shared by the target word and the pseudoword mask. When homophonic (MAYD) and orthographically similar (MARD) masks were equated for number of letters shared with the word target (made), both conditions led to a higher percentage of correct identifications of the target than a control mask, but an additional improvement in performance was found for the homophonic mask over the orthographically similar mask. The authors ascribed this effect to "phonetic activation" and concluded that such activation occurred automatically (nonoptionally).

They further argued that the effects arose before word recognition, assuming that the process of target identification, still in progress at the onset of the mask, could be influenced by the mask's orthographic and phonological properties. In the authors' view the extent to which different types of information (graphemic, phonemic, categorical) contributed to the identification of a word would depend on the exact timing of activation patterns. Critical for the interpretation of visual word recognition

---

1 The term "phonetic" is often used to indicate physical characteristics of the speech signal. We therefore prefer to use the term "phonological" to refer to abstract representations in the auditory modality in general, and the term "phonemic" when phoneme representations are involved. In the same fashion we use the terms "orthographic" and "graphemic" when talking about abstract representations in the visual domain.
experiments that manipulate the orthographic and phonological similarity of word and nonword stimuli is, therefore, the disentanglement of lexically and nonlexically mediated phonological effects, and of orthographic and phonological factors.

In contrast, bimodal studies have demonstrated that it is possible to test whether orthographic representations indeed activate associated phonological representations without these various confounding factors (e.g., Hanson, 1981; Kirsner, Milech, and Standen, 1983). In one such study, Frost, Repp, and Katz (1988) had subjects detect the presence of speech (either words or nonwords) in an auditory signal consisting of speech plus noise or noise alone. Simultaneous with the onset of the auditory stimulus, a matching visual stimulus (word or regular nonword), a nonmatching stimulus (word or nonword with similar structure), or a neutral stimulus (a row of Xs) appeared. When amplitude-modulated masking noise was used, matching visual stimuli biased the subjects to report that speech was presented but did not improve the detectability of speech in noise. However, the RTs for correct word detections were facilitated with respect to the neutral and the no-match conditions. In an experiment involving nonwords much smaller bias effects were found, and RTs for correct detections were not faster with a visual-auditory match. On the basis of these results, the authors suggested that printed words were immediately and automatically recoded into an internal phonetic form. Since the effects for nonwords were much weaker, they proposed that the influence of the visual stimulus on speech processing was lexically mediated.

This last conclusion is at variance with the results of the visual experiments reported above that suggest cross-modal activation occurs before word recognition (cf. also Gordon and Meyer, 1984, Experiments 1 and 5). Possibly this specific task, involving only the detection of speech, was not sensitive enough to pick up such influences. To better understand cross-modal sublexical activation, we examined the influence of matching and non-matching visual context on speech processing with a bimodal task that required not just detection but identification of the auditory stimulus. By varying not only the structural but also the temporal relationship between the visual and auditory stimuli we hoped to find out how and when representations of letters influence those of speech sounds.

More specifically, Dutch subjects made a forced two-choice decision on the identity of target vowels (e.g., /a:/ or /e:/) in auditory syllables (e.g., /ta:/ or /ke:/). Before, at, or after the onset of the auditory syllable, visual letter primes of two types could appear: indirect or direct primes. In indirect priming the letter was nominally congruent or incongruent with the consonant appearing in the same syllable as the target vowel (e.g., letter P, syllable /pa:/ with target vowel /a:/). Indirect priming effects on the vowel would indicate that cross-modal influences occur even when the visual stimulus is neither directly task-relevant nor connected to a response. In other words, such effects could be

---

2 The term “sublexical” is used here to indicate any representation smaller than the word. Such a representation may be called “prelexical” if it is computed as an intermediary representation before word recognition takes place. In this paper the term “lexical” is used to refer to word representations (e.g., “lexically mediated” is taken to mean “mediated by word representations”).
considered “automatic” (Posner and Snyder, 1975). In contrast to indirect priming in
which the letter was congruent or incongruent with the consonant in the target syllable,
in direct priming the letter was congruent or incongruent with the target vowel itself (e.g., letter A with syllable /a:/, where /a:/ indicates pronunciation of the letter A in Dutch). Once the existence of a connection between representations for a letter and a corresponding target vowel is shown, this direct priming technique should allow us to study the temporal development of the cross-modal influence. In sum, we first investigated the existence of automatic cross-modal effects in two experiments using the indirect form priming technique. We subsequently explored temporal aspects of cross-modal influences in more depth in a third experiment involving direct as well as indirect priming.

EXPERIMENT 1

In Experiment 1 the visual stimulus was always a letter that was either name-congruent or -incongruent with the consonant of an auditory CV-syllable (e.g., letter P or K, syllable /pa/). We expected that the congruence or incongruence between the letter and consonant would influence the response to the target vowel. This would be the case if the following assumptions hold. First, a graphemic representation of a letter is connected to its corresponding phonemic representation. Second, activation of the graphemic representation spreads to the corresponding phoneme. Third, more efficient processing of the consonant in the auditory CV-syllable leads to faster identification of the vowel in that syllable. The first two assumptions are compatible with several connectionist models of word recognition (e.g., variants of the time-course model of visual word recognition by Seidenberg et al., 1984). We expected the third assumption to hold on the basis of results like those of Wood and Day (1975) showing that the phonetic information corresponding to the consonantal and vocalic parts of such CV-syllables is processed as an integral unit. These authors showed that speeded phonetic decisions about vowels were influenced by the preceding stop (and vice-versa). In our experimental situation, the plosive consonant should be activated by the congruent letter, but not by the incongruent one; the more activated consonant should lead to faster identification of the following vowel. Hence, a facilitatory influence of the visual letter information upon the processing of the congruent consonant should also be observed upon the processing of the vowel.3

In sum, these three assumptions led us to predict faster RTs in a consonant-congruent condition (like P-/pa/) than in a consonant-incongruent condition (like K-/pa/). Further-

---

3 Another mechanism that does not assume that consonants and vowels are processed as integral units can be found in the TRACE I model of auditory word recognition (Elman and McClelland, 1986). In this model units representing stops are allowed to modify connections between units for features and vowel units that follow. Thus, a consonant like [p] would change the connection between a vowel like [a] and certain features of it (e.g., formant values), facilitating the recognition of the vowel by compensating for the contextual effects observed in the signal.
more, by varying the stimulus onset asynchrony (SOA) of the prime and the target syllable, we hoped to determine the time range over which the expected cross-modal influence could be obtained. Finally, we reasoned that any results showing priming should reflect automatic cross-modal activation, since in the indirect priming procedure used, the letter's identity is irrelevant to the forced-choice response to the vowel and consequently is unlikely to induce conscious strategies in the subject.

Method

Subjects. Thirty undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.

Design and stimuli. Bimodal test stimuli were used consisting of a visually presented letter and an auditorily presented nonword syllable. The letter was P, T, or K. The six different syllables consisted of a consonant (/p/, /t/, or /k/) and a short vowel (/a/ or /o/). All letters were combined with all syllables, leading to 18 different bimodal test stimuli. Depending on the relation between the letter and the syllable, a bimodal stimulus belonged to one of two experimental conditions: a consonant-congruent condition, in which the letter was nominally identical to the consonant in the auditory syllable (e.g., letter P paired with syllable /pa/); or a consonant-incongruent condition in which the letter and the consonant differed (e.g., letter K and syllable /pa/).

The temporal relationship between the visual and auditory stimuli was varied: The stimulus onset asynchrony (SOA) was –190, –70, –30, +30, or +150 msec. For example, at SOA1 the visual stimulus was presented 190 msec before the onset of the auditory syllable, while at SOA5 the visual stimulus was presented 150 msec after the onset of the auditory syllable.

The six stimuli in the congruent condition were repeated 10 times under each SOA, while the 12 stimuli in the incongruent condition were repeated five times under each of the five SOAs. To keep the subject’s attention directed to the screen, 80 catch trials (in which the subjects should not respond) were constructed in which the visual letter H was combined with all target syllables under the first three SOAs.

The resulting 6 × 10 × 5 (congruent) + 12 × 5 × 5 (incongruent) + 80 (catch) = 680 experimental trials were randomized and divided into two blocks of 340 trials each. Furthermore, 20 practice trials were added, bringing the total number of trials in the experiment to 700.

The auditory stimuli were recorded on tape by a female native speaker of Dutch in a sound-proof room. The length of the stimuli varied from 180 to 200 msec. The stimuli were digitized on a VAX–11/750 computer with a sampling rate of 20 kHz, and a randomized sequence of targets was placed on one channel of a tape. The output of the computer was low-pass filtered with a cutoff frequency of 10 kHz. During the experiment the auditory stimuli were presented binaurally over headphones. On the second channel of the tape a pulse, inaudible to the subjects, was placed that triggered both the timer for the recording of the response latencies and the presentation of the visual stimulus (after a delay that depended on the specific SOA in a trial).

The visual stimuli were white Roman capitals, 6 mm in height, presented on a MATROX-screen with a dark background. The monitor was placed at a distance of 60
cm from the subject, in order to provide projection within the foveal field of the eye. The visual stimuli were presented for 60 msec. Presentation of the visual stimuli and recording of the reaction times were controlled by a PDP-11/23 computer.

Procedure. Subjects read the written instructions, which were repeated orally at the beginning of the experiment. They were instructed to rest their two index fingers lightly on the two response buttons in front of them, and to push the /a/-button as fast as possible whenever they heard the /a/-vowel, and push the /o/-button when they heard the /o/-vowel. The /o/-response button was allocated to the index finger of the preferred hand. Subjects were also told that sometime before or during the presentation of the auditory syllable, a letter would appear on the screen, and that they were not to respond if the letter was H.

Each trial consisted of a 1000 Hz warning signal of 150 msec duration, followed after 500 msec of silence by the auditory syllable. A variable interval after the warning signal (depending on the SOA in the trial) a visual letter stimulus appeared. A new trial was initiated every 5.5 seconds.

The session consisted of 20 practice trials followed by two blocks of 340 test trials. After the practice set there was a short pause in which subjects could ask for clarifications. Between the two experimental blocks there was a pause of three minutes. In all, the experiment took about one hour and fifteen minutes.

Results

Mean RTs (measured from the onset of the auditory syllable) were computed for each subject and for congruent and incongruent conditions at each SOA. The percentage of missing RTs and RTs greater than 1000 msec or smaller than 150 msec was 6.5, and was distributed equally across the congruent and incongruent conditions. Missing values were substituted by mean RTs in each relevant subcondition for each subject. The percentage of errors on the H-trials (false alarms) was 3.6. Table 1 shows the mean RTs for congruent and incongruent conditions under each of the five SOAs.

An analysis of variance with the factors Congruence (congruent vs. incongruent), Syllable Type (with initial consonants /p/, /t/, or /k/) and SOA showed significant main effects for all these factors (Congruence [F(1, 29) = 26.80, p < 0.001]; Syllable Type [F(2, 58) = 11.42, p < 0.001]; and SOA [F(4, 116) = 40.96, p < 0.001]. A significant interaction was found between Congruence and Syllable Type [F(2, 58) = 5.71, p < 0.01]. No significant interactions of any factors with SOA were obtained.

Paired planned comparisons of the congruent and the incongruent conditions for each of the three syllable types showed significant differences for syllables with /p/- and /t/-consonants, but not for those with /k/-consonants, as indicated in Table 2. A further analysis of the syllables containing a /k/-consonant showed that the RT-difference between congruent and incongruent conditions was significant for the syllable /ka/ (535 vs. 548 msec, t(29) = 2.68, p = 0.01) but not for /ko/ (563 vs. 557 msec, t(29) = 0.98, p > 0.10).
Table 1

Mean reaction times (in msec) for consonant-congruent and -incongruent conditions as a function of SOA

<table>
<thead>
<tr>
<th>condition</th>
<th>SOA</th>
<th>-190</th>
<th>-70</th>
<th>-30</th>
<th>+30</th>
<th>+150</th>
</tr>
</thead>
<tbody>
<tr>
<td>consonant-congruent</td>
<td>504</td>
<td>529</td>
<td>541</td>
<td>549</td>
<td>571</td>
<td></td>
</tr>
<tr>
<td>consonant-incongruent</td>
<td>521</td>
<td>541</td>
<td>549</td>
<td>564</td>
<td>582</td>
<td></td>
</tr>
</tbody>
</table>

Table 2

Mean reaction times (in msec) and planned comparisons of means for /p/-, /t/-, and /k/-syllable types in consonant-congruent and -incongruent conditions

<table>
<thead>
<tr>
<th>condition</th>
<th>congruent</th>
<th>incongruent</th>
<th>t(29)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>/p/-syllable</td>
<td>539</td>
<td>554</td>
<td>4.61</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>/t/-syllable</td>
<td>528</td>
<td>547</td>
<td>4.64</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>/k/-syllable</td>
<td>549</td>
<td>552</td>
<td>0.95</td>
<td>ns</td>
</tr>
</tbody>
</table>

Discussion

The most important finding of this experiment was a significant RT-advantage of the consonant-congruent condition over the -incongruent condition. This congruence effect can best be understood as being the result of cross-modal activation of an auditory consonant representation by a visual letter representation. To obtain these results in an indirect priming procedure where grapheme-phoneme congruence by itself has no predictive value for the vowel response suggests that the cross-modal activation is automatic. Indeed, for choosing the correct vowel a cross-modal comparison is neither required nor directly relevant. Consequently, these results provide evidence for automatic activation of phoneme representations by graphemes.

The absence of an interaction between the factors, Congruence and SOA, raises some questions about the temporal characteristics of this cross-modal activation of the consonant by the letter prime. One might have expected the influence of the letter prime upon the processing of the vowel to decrease or even to disappear at later SOAs, but it appears that this facilitatory influence remained roughly constant across the entire SOA range. However, there was still an interval of approximately 400 msec between the
onset of the letter (in the latest SOA-condition) and the response, during which the
cross-modal effect could have taken place.

The observed gradual increase in RT over SOA for consonant-congruent and
-incongruent conditions could be a consequence of the catch-trial condition with H. In
order to avoid reacting in H-trials, subjects might delay their response until the letter
is at least partially identified. Thus, the later the letter appears in time relative to the
auditory syllable, the longer the subject would have to wait. This potentially negative
consequence of the H-condition led us to replicate the experiment with some changes
in design.

**EXPERIMENT 2**

In Experiment 2, the catch-trial condition was eliminated and the subject only made a
decision on the vowel in the syllable. To ensure that the subject paid attention to the
visual stimuli, an additional off-line task was used: After a varying number of trials, the
word “RAPPORT” (“report”) appeared, indicating that the subject had to write down
whether a letter was presented on the last trial or not. To make this off-line decision
non-trivial, Experiment 2 included auditory single-channel trials. This single-channel
condition served also as a baseline for the evaluation of congruence effects. Finally,
we attempted to determine whether the differences between consonant-congruent and
-incongruent conditions generalized to other syllable types. For this purpose, fricative–
vowel syllables were included, as it has been shown that the cues to their component
phonemes are processed interdependently (Tomiak, Mullennix, and Sawusch, 1987;
Whalen, 1984).

**Method**

**Subjects.** Twenty-seven undergraduates at Nijmegen University, all native speakers of
Dutch, were paid to participate in this experiment.

**Design and stimuli.** The design of this experiment was similar to that of Experiment 1.
However, in addition to the two experimental conditions of Experiment 1, a single-
channel condition was included. Bimodal stimuli consisted of a visually presented letter
(P, K, or S) and an auditorily presented CV-syllable. The consonant of the syllable
was /p/, /k/, or /s/; the vowel was /a:/ or /e:/; The choice of long vowels in this
experiment was motivated by their predominance in free-occurring Dutch CV-syllables.  

4 In a pilot run we had subjects identify the letter that appeared in the preceding trial.
Since they kept confusing the presented letter with the auditory target syllable, in
the experiment we only asked whether any visual stimulus had been presented at all.
Still, the confusion observed already seems to indicate that some kind of automatic
cross-modal activation was going on.

5 The vowel /e:/ was chosen for two main reasons. First, we hoped to avoid deviating
results like that for the syllable /ko/ in Experiment 1. Second, Experiments 2 and 3
also involved a manipulation of letter name vs. abstract letter representation. E.g.,
The length of the six syllables varied from 430 to 500 msec. The single-channel conditions consisted of syllable presentation without a letter. All visual stimuli were presented for 100 msec.

Stimuli in the congruent condition (e.g., letter S combined with syllable /se:/) were repeated 10 times under each of five SOAs, while stimuli in the incongruent condition (e.g., letter K with syllable /sa:/) were repeated five times under each SOA. The SOAs were −190, −70, −30, +30, and +150 msec. In addition, there were 30 single-channel trials of each of the six auditory syllables. Thus, the experiment consisted of $6 \times 10 \times 5$ (congruent) + $12 \times 5 \times 5$ (incongruent) + 180 (single-channel) = 780 test stimuli. These were randomized and divided into two sessions of 390 trials each. Furthermore, for each session 60 report trials, which consisted of the visual presentation of the word ‘RAPPORT’, and 47 practice trials were constructed. In all, each experimental session included 497 trials.

Procedure. The same two-choice procedure was used as in the previous experiment. Subjects participated in the two sessions on successive days. The order of sessions was counterbalanced over subjects. Response allocation across the two sessions was also counterbalanced over subjects. Half of the subjects reacted to /e:/ with their right index finger on the right response button in the first session, and with the left index finger on the left response button in their second session the next day. The other half of the subjects was instructed to do just the opposite. In addition, whenever the word ‘RAPPORT’ was visually presented, subjects were to indicate on a prestructured form whether the preceding trial had included a visual stimulus or not. A four second interval followed each trial, except report trials, which were followed by a pause of six seconds. Each experimental session lasted for 40 minutes with a short break after about 20 minutes.

Results

Mean RTs (measured from the onset of the auditory syllable) were computed for each subject and each experimental condition. The percentage of missing RTs and RTs greater than 1000 msec or smaller than 150 msec was 4.3, and was distributed equally across consonant-congruent and -incongruent conditions. Missing values were substituted by mean RTs in each subcondition for each subject. The percentage of errors in the subjects’ judgement about whether a letter appeared in the trial preceding the RAPPORT-trials was 5.6. Mean RTs per SOA for the congruent and incongruent conditions as well as for the single-channel condition are shown in Table 3.

An analysis of variance with the factors Congruence (consonant-congruent vs. -incongruent), SOA, and Syllable Type (with initial consonants /p/, /k/, or /s/) showed

in Dutch the letter name for P is /pe:/ and for K is /ka:/ By using target syllables like /pa:/, /pe:/ and /ka:/, /ke:/ the effect of letter name vs. abstract consonant representation on the auditory syllable could be distinguished. Since this manipulation did not lead to consistent results over syllables and experiments, the effects of this factor were not further documented.
TABLE 3

Mean reaction times (in msec) for the single-channel condition, and for the consonant-congruent and consonant-incongruent conditions as a function of SOA

<table>
<thead>
<tr>
<th>condition</th>
<th>SOA</th>
<th>-190</th>
<th>-70</th>
<th>-30</th>
<th>+30</th>
<th>+150</th>
</tr>
</thead>
<tbody>
<tr>
<td>consonant-congruent</td>
<td>509</td>
<td>541</td>
<td>540</td>
<td>552</td>
<td>561</td>
<td></td>
</tr>
<tr>
<td>consonant-incongruent</td>
<td>525</td>
<td>557</td>
<td>553</td>
<td>571</td>
<td>571</td>
<td></td>
</tr>
<tr>
<td>single-channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>558</td>
</tr>
</tbody>
</table>

main effects of Congruence [F (1, 26) = 39.82, p < 0.001], SOA [F (4, 104) = 42.06, p < 0.0001], and Syllable Type [F (2, 52) = 95.50, p < 0.0001]. A significant interaction was found between Congruence, SOA and Syllable Type [F (8, 208) = 3.83, p < 0.001]. The other interactions were not significant.

Discussion

Both Experiments 1 and 2 showed faster RTs in the consonant-congruent than in the consonant-incongruent conditions, despite several differences in the procedure used. Congruence effects were found for syllables with stop consonants (Experiments 1 and 2) and with fricative consonants (Experiment 2); and when the task explicitly demanded on-line identification of the visual stimulus (Experiment 1), and when it did not (Experiment 2). Furthermore, both experiments showed a congruence effect over the entire SOA range from −190 (letter first) to +150 (syllable first), showing that even when the auditory syllable leads by 150 msec, response time to the vowel is influenced by the congruence or incongruence of letter and consonant.

The existence of this congruence effect provides strong evidence for sublexical activation of auditory consonant representations by congruent visual letters. Since the relationship between the auditory consonant and the letter is not relevant for the subject’s vowel decision, we can conclude that the activation takes place automatically, that is, without conscious control by the subject (Posner and Snyder, 1975). This demonstration of cross-modal activation with the indirect priming procedure confirms and extends the results of other studies such as those using bimodal same/different matching (cf. Wood, 1977; Posner, 1978), where a direct comparison between graphemes and phonemes is required by the task itself.

Experiments 1 and 2 both revealed a gradual increase in the RTs for all conditions with increasing SOA. One explanation proposed for this effect in Experiment 1 depended upon the presence of the catch-trials. It was argued that subjects delayed their response until they had partially identified the visual stimulus, because they were not to respond when the letter H appeared. However, since Experiment 2 showed an increase of RT
over SOA even without this catch-trial condition, we can eliminate this explanation.

According to another explanation for the increase of RT over SOA, the arrival of the visual stimulus causes a shift of attention away from the auditory modality and disrupts processing (cf. Miller, 1985, p. 520). If this attention shift occurs while the auditory stimulus is being processed (late letter presentation), it should lead to greater disruptions and stronger increases in RT than when the auditory stimulus has not yet arrived (early letter presentation). This view, in which the visual stimulus has a detrimental effect, may be contrasted with one in which the visual stimulus functions as an alerting cue, leading to preparation-enhancement (cf. Nickerson, 1973). The amount of facilitation caused by the visual stimulus would depend on its relative time of arrival with respect to the auditory target stimulus: Early visual stimuli would facilitate the RT more than late. If the single-channel condition is taken as a baseline, the increase in RT over SOAs is better explained by the preparation-enhancement hypothesis, since for most SOAs the bimodal conditions show a facilitatory effect (see Table 3).

EXPERIMENT 3

The previous experiments using the indirect priming technique have not allowed us to determine the time-course of phoneme activation by graphemes, since there was no interaction between the size of the congruence effect and the SOA. Furthermore, the interpretation of the results in terms of the timing of the cross-modal influence is complicated here because it is difficult to untangle the temporal properties of this influence from those of the processing of the vowel. As a consequence, we decided to prime the auditory vowel directly in order to obtain a more direct reflection of the spread of activation from the visual prime to the auditory target under different SOAs. In the next experiment we therefore added conditions in which the letter prime was either congruent with the target vowel or not (e.g., in Dutch, A-/ka:/ vs. A-/ke:/).

As a baseline condition for each SOA, a bimodal condition was included in which a nonlinguistic visual stimulus accompanied the auditory syllable. Effects of attention shift or alerting caused by the presence of a visual stimulus should be similar for baseline and test conditions. As in Experiment 2, an auditory single-channel condition was also included to examine the general effect of adding a visual accessory to the auditory syllable.

In addition to CV-syllables (like /ka:/), Experiment 3 also included syllables with two other structures: VC-syllables (like /a:k/) and V-syllables (like /æ:/). If the integration account described above is correct, subjects would integrate the consonantal information with that of the target vowel in the case of VC-syllables. Thus, the RTs to VC-syllables should be slower than those to V-syllables, provided that vowel processing is not complete when the consonant of a VC-syllable arrives.

Method

Subjects. Twenty-eight undergraduates at Nijmegen University, all native speakers of Dutch, were paid to participate in the experiment.
Design and stimuli. Visually presented stimuli were the letters K, P, A, and E and the symbol *; auditory stimuli were the syllables /ka:/, /ke:/, /a:/, /e:/, /a:k/, and /e:k/.

The naturally pronounced syllables were matched in length as much as possible: The duration of /ka:/', /ke:/', /a:/', and /e:/ was about 365 msec; the duration of /a:k/ was about 465 msec, that of /e:k/ 495 msec. The length of the vowel in /a:k/ and /e:k/ was about 215 msec. All visual stimuli were presented for 100 msec.

Each visual stimulus was combined with each syllable. In the consonant-congruent conditions the letter was nominally identical to the consonant (letter K combined with syllables /ka:/, /ke:/, /a:k/, or /e:k/, in the consonant-incongruent conditions it was not (letter P combined with these syllables). In the vowel-congruent conditions the letter was nominally congruent to the target vowel (e.g., letter A combined with /ka:/, /a:/', or /a:k/), in the vowel-incongruent conditions it was not (e.g., letter E combined with syllables just mentioned). Furthermore, there was a condition in which a letter consonant was combined with a V-syllable (e.g., P/e:/). In the star-condition the auditory syllable was presented with a visual *. A single-channel condition was also included, in which the syllable was presented in isolation.

As before, the temporal relationship between the visual and auditory stimulus in bimodal conditions was varied. However, in order to examine possible cross-modal effects over a longer time range, the SOAs used in this experiment were −250, −100, 0, +100, and +250 msec. A negative number indicates the visual context preceded the auditory syllable.

Each bimodal stimulus was repeated 10 times under each SOA. Each single-channel stimulus was repeated 20 times. Therefore, the number of test stimuli in the experiment was 6 (syllables) X 5 (visual stimuli) X 5 (SOAs) X 10 (repetitions) + 6 (syllables) X 20 (repetitions) = 1620. These were randomized and divided into three experimental sessions such that each session consisted of 540 trials. Furthermore, 60 report trials (consisting of the visual presentation of the word ‘RAPPORT’) and 40 practice trials were constructed for each session. Thus, subjects were presented with 640 trials per session.

Procedure. Subjects participated in three sessions on successive days. The order of sessions was counterbalanced over subjects. Half of the subjects reacted to /e:/ with their right index finger, and to /a:/ with the left index finger. The other half of the subjects was instructed to do just the opposite. Thus handedness and response button were counterbalanced. In other respects the procedure was the same as in the previous experiment. Each experimental session lasted for about an hour and had a short break after about 35 minutes.

Results

Mean RTs were computed for each subject and each experimental condition. The

---

6 While the syllables in Experiment 1 were all clearly nonwords, some of those in Experiment 2 could in fact be treated as words by the subjects (e.g., /pa:/ and /te:/). Since many CV-syllables have lexical status, this is hard to avoid. Because the syllables containing a /k/-consonant were all either nonwords or words of a negligible frequency, this consonant was chosen for the CV- and VC-syllables in Experiment 3. Furthermore, no clear RT-differences were found in Experiments 2 and 3 between syllables that could be considered words or not.
TABLE 4

Mean reaction times (in msec) for the single-channel condition, and for the consonant-congruent, -incongruent, and star-conditions as a function of SOA

<table>
<thead>
<tr>
<th>condition</th>
<th>SOA</th>
<th>-250</th>
<th>-100</th>
<th>0</th>
<th>+100</th>
<th>+250</th>
</tr>
</thead>
<tbody>
<tr>
<td>consonant-congruent</td>
<td>415</td>
<td>428</td>
<td>433</td>
<td>426</td>
<td>432</td>
<td></td>
</tr>
<tr>
<td>consonant-incongruent</td>
<td>419</td>
<td>423</td>
<td>436</td>
<td>426</td>
<td>443</td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>421</td>
<td>424</td>
<td>435</td>
<td>429</td>
<td>440</td>
<td></td>
</tr>
<tr>
<td>single-channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>444</td>
</tr>
</tbody>
</table>

TABLE 5

Mean reaction times (in msec) for the single-channel condition, and for the vowel-congruent, -incongruent and star-conditions as a function of SOA

<table>
<thead>
<tr>
<th>condition</th>
<th>SOA</th>
<th>-250</th>
<th>-100</th>
<th>0</th>
<th>+100</th>
<th>+250</th>
</tr>
</thead>
<tbody>
<tr>
<td>vowel-congruent</td>
<td>384</td>
<td>385</td>
<td>411</td>
<td>424</td>
<td>433</td>
<td></td>
</tr>
<tr>
<td>vowel-incongruent</td>
<td>433</td>
<td>432</td>
<td>449</td>
<td>424</td>
<td>431</td>
<td></td>
</tr>
<tr>
<td>star</td>
<td>409</td>
<td>415</td>
<td>428</td>
<td>424</td>
<td>433</td>
<td></td>
</tr>
<tr>
<td>single-channel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>437</td>
</tr>
</tbody>
</table>

The percentage of missing RTs and RTs greater than 1000 msec or smaller than 150 msec was 1.2 for the double channel conditions and 2 for the single-channel condition. Missing values were substituted by mean RTs in each condition. The percentage of errors in the subjects’ response on the RAPPORT-trials was 6.0.

The data were analysed separately for the two types of congruence: First results for all consonant-congruent and -incongruent conditions involving CV- and VC-syllables are presented, then results for the vowel-congruent and -incongruent conditions. The corresponding star-conditions were included in the analyses.

Mean RTs are given for the Consonant-Congruence data in Table 4. An analysis of variance on these data with the factors Subjects, Congruence Type (consonant-congruent, consonant-incongruent, star), and SOA led to a main effect for SOA \(F(4, 108) = 7.67,\)
Table 6
Mean reaction times (in msec) for CV-, V-, and VC-syllable types in the single-channel condition, and in the combined vowel-congruent, -incongruent and star-conditions as a function of SOA

<table>
<thead>
<tr>
<th>condition</th>
<th>SOA</th>
<th>-250</th>
<th>-100</th>
<th>0</th>
<th>+100</th>
<th>+250</th>
<th>single-channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV-syllable</td>
<td>443</td>
<td>439</td>
<td>456</td>
<td>446</td>
<td>453</td>
<td>460</td>
<td></td>
</tr>
<tr>
<td>VC-syllable</td>
<td>394</td>
<td>401</td>
<td>418</td>
<td>415</td>
<td>428</td>
<td>429</td>
<td></td>
</tr>
<tr>
<td>V-syllable</td>
<td>389</td>
<td>392</td>
<td>413</td>
<td>410</td>
<td>417</td>
<td>423</td>
<td></td>
</tr>
</tbody>
</table>

$p < 0.0001$. No significant main effect was found for Congruence Type [$F < 1$], nor any significant interaction of Congruence Type and SOA [$F < 1$].

Mean RTs are given for the Vowel-Congruence data in Table 5. An analysis of variance on these data with the factors Subjects, Congruence Type (vowel-congruent, vowel-incongruent, star) and SOA showed significant main effects for Congruence Type [$F (2, 54) = 49.34, p < 0.0001$], and SOA [$F (4, 108) = 19.84, p < 0.0001$], and a significant interaction effect for Congruence Type and SOA [$F (8, 216) = 16.91, p < 0.0001$].

We further related the vowel-congruence and -incongruence conditions to the star- and single-channel conditions. Planned comparisons between the star-conditions and the vowel-congruent or -incongruent conditions for each SOA showed significant differences for all comparisons under the first three SOAs (cf. Table 5). Planned comparisons between all these bimodal conditions and the single-channel condition showed significant differences for all comparisons under these SOAs as well, except for the vowel-incongruent condition, that did not differ significantly from the single-channel condition at SOA1 and SOA 2 (Table 5). At SOA4 all bimodal conditions were significantly faster than the single-channel condition, but did not differ among each other. At SOA5 no significant differences were found any more.

In order to test if the three Syllable Categories (CV, V, and VC) behaved differently over SOA, an analysis of variance was run on the combined vowel-congruence data (vowel-congruent, vowel-incongruent, and star-conditions) with the factors Subjects, Syllable Category, and SOA. Mean RTs, including those of the single-channel condition, are given in Table 6. The analysis showed significant main effects for Syllable Category [$F (2, 54) = 95.75, p < 0.0001$], SOA [$F (4, 104) = 19.84, p < 0.0001$], and a significant interaction of Syllable Category and SOA [$F (8, 216) = 2.85, p < 0.01$].

Planned comparisons between the different Syllable Categories for each SOA showed that reactions were significantly faster to V-syllables than to CV-syllables at all five SOAs, but faster than to VC-syllables at only two SOAs (SOA2 = -100 and SOA5 = +250).
Discussion

Experiments 1 and 2 have demonstrated automatic cross-modal grapheme-to-phoneme activation. Experiment 3 provides information on how such activation develops across time. Large differences between vowel-congruent and -incongruent conditions were found when the letter appeared simultaneously with or before the auditory syllable. In order to distinguish facilitatory and inhibitory components of this vowel-congruence effect, these conditions were compared to the star-condition. Relative to this baseline, strong facilitation effects were obtained for the vowel-congruent condition and strong inhibition effects for the -incongruent condition, when the visual stimulus preceded the auditory syllable by 250 or 100 msec. Both effects disappeared completely when the letter was presented 100 or 250 msec after the onset of the syllable. The facilitation effect can be interpreted as evidence for cross-modal activation of the target vowel-representation by the letter stimulus, while the inhibition effect suggests cross-modal activation of the phonological representation of the vowel associated with the competing response.

Comparison of the data with the single-channel condition revealed a more global facilitatory influence of the visual stimulus. Even when the auditory syllable led by 100 msec (SOA4), and the RT-differences between vowel-congruent, -incongruent and star-conditions had disappeared, a significant facilitation remained in comparison with the single-channel condition. This effect disappeared only at an SOA of 250 msec. We attribute this effect to a general non-specific alerting of the subject by the visual stimulus, as we did for the decrease in RTs for the negative SOAs in the first two experiments.

The analysis of the RTs for the different syllable categories (CV, V, or VC) over the combined vowel-congruent, -incongruent, and star-conditions showed that RTs to CV-syllables were slower than those to V- and VC-syllables by 25 to 50 msec. Since RTs were measured from stimulus onset, this difference is best accounted for by the delayed arrival of the vowel in the CV-syllables compared to the V- or VC-syllables. Slightly slower RTs were found to VC-syllables than to V-syllables, indicating that the subjects took into consideration some information about the following consonant as was predicted by the account appealing to the integration of consonant and vowel information (cf. Healy and Cutting, 1976).

The RT-differences between consonant-congruent and -incongruent conditions, robust in Experiments 1 and 2, disappeared in Experiment 3. To the extent that the literature on semantic priming is relevant here, we can find some explanations for this unexpected result. For example, it is known that the types of experimental conditions, as well as their relative proportion, influence the size of the semantic priming effects observed (e.g., Neely, 1977). The inclusion of the vowel-congruent and vowel-incongruent conditions in our Experiment 3 changed the proportion of consonant-congruent conditions, thus perhaps washing out the more subtle effects for the consonant-congruent and -incongruent conditions. Furthermore, the priming literature suggests that faster RTs generally lead to smaller priming effects (e.g., Flores d’Arcais, Schreuder, and Glazeborg, 1985). The fact that the RTs in Experiment 3 were about 100 msec faster than those in the first two experiments may have contributed to the disappearance of the consonant-congruence effect (cf. the range effects described by Poulton, 1973). Further experimentation is
clearly needed to investigate how appropriate these explanations really are for our form priming experiments.

**GENERAL DISCUSSION**

In a series of three experiments we examined the cross-modal influence of graphemes on phonemes and the time-course of this influence. The existence of cross-modal activation was demonstrated in Experiments 1 and 2 using an indirect priming technique in which the priming visual letter stimuli were either congruent or incongruent with the consonant of an auditory syllable (e.g., letter P or K, syllable /pa:/). Over the entire range of SOAs used, an RT-advantage for the vowel target was found in the consonant-congruent over the -incongruent condition. To explain this advantage, we must assume automatic cross-modal activation of the representation of the consonant phoneme by the letter prime.

The temporal development of cross-modal sublexical activation was explored in more depth in Experiment 3 with a direct priming technique. Here the letter stimuli presented were either congruent (e.g., letter A, syllable /ka:/) or incongruent (letter E, syllable /ka:/) with the target vowel in the auditory syllable. For trials in which the letter preceded the auditory syllable, strong facilitation and inhibition effects were obtained when the vowel-congruent and -incongruent conditions were compared to the bimodal baseline-condition. These effects decreased when the letter was presented at the onset of the auditory syllable, and disappeared when the letter was presented during the auditory stimulus. The evolution of the facilitatory and inhibitory effects over SOA suggests that it took time for the letter to activate its corresponding phonemic representation. When the letter came late, its contribution to the activation of this phonemic representation could no longer influence the response.

More specifically, the results of Experiment 3 support the following account of bi-modal processing in this forced-choice task. In the bimodal conditions, the letter accompanying the auditory signal generates an orthographic representation which automatically activates an associated phonological representation. In the vowel-congruent condition with negative SOAs, the letter preactivates the same representation as is activated by the vowel, and this additional activation speeds up the response compared to the base-line condition. Indeed, in this baseline condition, the star symbol does not activate any phonological representation. When the phonological representation activated by the letter differs from that of the target vowel, its influence depends on whether it belongs to the response set or not. When it does not belong to the response set (e.g., the phonological representation corresponding to the letter P), it does not influence the response, producing no RT-difference between this condition and the baseline star-condition. However, when the phonological representation activated by the letter (e.g., E) does belong to the response set, the target vowel and the letter activate phonological representations associated with competing responses, leading to the inhibition found for the vowel-incongruent condition.

The explanation we have advanced for our results is similar to that proposed for the results obtained with a family of tasks often applied in the domain of visual word
Grapheme Effects on Phonemic Processing

recognition, such as the Stroop task (e.g., Stroop, 1935; Glaser and Glaser, 1982), the picture-word naming task (e.g., Glaser and Düngelhoff, 1984; LaHeij, 1988) and the flanker task (e.g., Eriksen and Eriksen, 1974; Hell, 1987). In a typical Stroop task, color words (like RED) are printed in a color that is congruent or incongruent with their name (e.g., in red or blue). If subjects have to name the color the words are printed in, naming time increases up to 100 msec when the print color and word meaning are incongruent, as compared to a situation where the color of a meaningless letter string is named. When this experimental situation is translated into that of Experiment 3, the target vowel corresponds to the color, and the visual letter to the orthographic form of the printed word.

As in the Stroop task, (vowel-)incongruent conditions led to interference effects, and (vowel-)congruent conditions to facilitation relative to a bimodal baseline condition. The development of the facilitation and inhibition effects over SOA is also comparable (cf. Glaser and Glaser, 1982; Glaser and Düngelhoff, 1984). With van der Heijden (1981) we would like to suggest that these similarities across tasks reflect general principles in the mechanisms used to perform different types of attentional tasks. Since Stroop effects are usually interpreted to be indicative of automatic processing (e.g., Kahneman and Chajczyk, 1983), the similarity of the current data and those obtained in Stroop tasks then implies the same for our congruence effects.

After having established the existence of automatic cross-modal activation of phonemes by graphemes, we must now assess the relevance of this conclusion for visual word recognition. If graphemic information automatically activates phonological representations in situations in which the visual stimulus serves mainly as an accessory and is not directly task-relevant, such cross-modal activation is even more likely to occur in visual word recognition, where attention must be paid to visual stimuli and where phonological information can play a useful role. We therefore would like to argue that our results, although obtained with nonword syllables, support the hypothesis that sublexical grapheme-to-phoneme activation occurs automatically in visual word recognition.

By measuring the activation of phonological representations by graphemes via an auditory task, we have not concerned ourselves directly with visual word recognition. Nonetheless, we have made certain that the representations activated are the same as those activated in auditory perception. This is not necessarily the case for visual tasks, since they do not directly measure phonological effects. There is no guarantee that the phonological effects arising for visually presented words or nonwords are truly "phonological". Thus, both approaches provide complementary evidence for grapheme-to-phoneme activation in word recognition. Taken together, the relevant experimental evidence obtained exclusively in the visual domain (Van Orden, 1987; Van Orden et al., 1988; Perfetti et al., 1988) and the bimodal results for words (e.g., Hanson, 1981) and nonwords (this work) offer convergent evidence for the existence of automatic, not lexically mediated, grapheme-to-phoneme activation during visual word recognition.

The activation metaphor we adopted in this paper seems particularly well-suited to account for the gradual decrease in facilitation and inhibition effects over SOA in Experiment 3. Indeed, all results are easily understood within the context of interactive activation models, such as the time-course model of visual word recognition (e.g., Seiden-
berg, 1985b). If the model were expanded to include the possibility of both orthographic and phonological input, the effects of visual context on auditory representations could be simulated. Furthermore, while our data are accounted for by this type of model quite well, they are hard to reconcile with all those models of visual word recognition in which phonological information becomes available only after word recognition.

In sum, bimodal experiments of the type presented here may help to constrain models of visual word recognition by providing information about structural and temporal aspects of cross-modal activation. Our results indicate that such models should incorporate a mechanism of sublexical activation of phonemic representations by graphemes.

(Received March 20, 1989; accepted August 11, 1989)

REFERENCES


Grapheme Effects on Phonemic Processing


