Article

Bidirectional grapheme-phoneme activation in a bimodal detection task

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

Abstract

A divided attention paradigm was used to investigate whether graphemes and phonemes can mutually activate or inhibit each other during bimodal processing. In 3 experiments, Dutch subjects reacted to visual and auditory targets in single-channel or bimodal stimuli. In some bimodal conditions, the visual and auditory targets were nominally identical or redundant (e.g., visual A and auditory /a/); in others they were not (e.g., visual U and auditory /a/). Temporal aspects of cross-modal activation were examined by varying the stimulus onset asynchrony of visual and auditory stimuli. Cross-modal facilitation--but not inhibition--occurred rapidly and automatically between phoneme and grapheme representations. Implications for current models of bimodal processing and word recognition are discussed.

Reference

Bidirectional Grapheme–Phoneme Activation in a Bimodal Detection Task

Ton Dijkstra, Uli H. Frauenfelder, and Robert Schreuder

A divided attention paradigm was used to investigate whether graphemes and phonemes can mutually activate or inhibit each other during bimodal processing. In 3 experiments, Dutch subjects reacted to visual and auditory targets in single-channel or bimodal stimuli. In some bimodal conditions, the visual and auditory targets were nominally identical or redundant (e.g., visual A and auditory /a/); in others they were not (e.g., visual U and auditory /a/). Temporal aspects of cross-modal activation were examined by varying the stimulus onset asynchrony of visual and auditory stimuli. Cross-modal facilitation—but not inhibition—occurred rapidly and automatically between phoneme and grapheme representations. Implications for current models of bimodal processing and word recognition are discussed.

"Assailed by eyes, ears, nose, skin, and entrails at once," we should experience the world "as one great blooming, buzzing confusion," as William James (1890/1950, p. 488) put it, were we not able to integrate and segregate information arriving simultaneously in different modalities as efficiently and rapidly as we do. For the visual and auditory modalities, experimental psychologists have studied these fast integration processes with both nonlinguistic and linguistic stimuli and at three general levels of processing: recognition, decision, and response (Miller, 1982). Using simple stimuli such as light flashes and beeps, researchers have examined how and when the processing of a visual signal can be influenced by an auditory signal, and vice versa. One tentative conclusion of this research has been that the processing of nonlinguistic stimuli can be influenced cross-modally at both response and decision levels (Craig, Colquhoun, & Corcoran, 1976; Diederich & Colonius, 1987; Gielen, Schmidt, & Van den Heuvel, 1983; Miller, 1982, 1985, 1986, 1991; Nickerson, 1973).

Ton Dijkstra, Nijmegen Institute for Cognition and Information, University of Nijmegen, and Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands; Uli H. Frauenfelder, Max-Planck-Institut für Psycholinguistik, The Netherlands; Robert Schreuder, Interfaculty Research Unit for Language and Speech, University of Nijmegen, The Netherlands.

Correspondence concerning this article should be addressed to Ton Dijkstra, Nijmegen Institute for Cognition and Information, University of Nijmegen, P.O. Box 9104, NL-6500 HE Nijmegen, The Netherlands. Electronic mail may be sent to dijkstra@nici.kun.nl.

When the structural and temporal aspects of the integration of linguistic stimuli in different modalities are considered, a whole set of additional research questions of interest arises because of the involvement of permanently stored linguistic representations. Mental representations at diverse linguistic levels are assumed to be involved in language processing. Of central interest for the present research was the nature of the interactions between phonological and graphemic representations during the word recognition process. Because these representations also become available postlexically after the retrieval of the lexical representations, it is important to ensure that the effects under scrutiny are due to prelexical cross-modal contacts between graphemes and phonemes.

Two complementary research strategies to this problem can be distinguished. Most commonly, grapheme-to-phoneme activation effects have been investigated by manipulating the orthographic and phonological properties of visually presented words. Indeed, there is a growing body of research on the role phonology plays in visual word recognition. On the basis of a comprehensive review of this literature, Van Orden, Pennington, and Stone (1990) rejected the hypothesis that the output of phonological mediation is delayed relative to the output of a direct visual route (the so-called delayed phonology hypothesis). Moreover, they also rejected a bypass hypothesis, which assumes that skilled readers can circumvent phonological mediation in visual word recognition through direct access. Van Orden et al. concluded that the phonology of visually presented words affects word identification within its normal time course.

Alternatively, cross-modal effects can be investigated with experimental techniques that present subjects with bimodal stimuli. These effects can be studied more directly with this approach than they can with pure visual word recognition experiments. The tacit assumption made in the latter studies that the phonological representations computed in visual processing are indeed the same as those involved in auditory processing can be avoided or even tested with a bimodal approach. More important, the use of bimodal stimuli allows the investigation of not only the effect of graphemes on phonemes, but also the inverse. In the present experiments, we exploited bimodal presentation techniques to study the relation between the processing of visually presented linguistic

stimuli and the processing of auditorily presented linguistic stimuli.

There are a few bimodal studies that have addressed cross-modal nonlexical activation (Dijkstra, 1990; Dijkstra, Schreuder, & Frauenfelder, 1989; Frost & Katz, 1989; Frost, Repp, & Katz, 1988). Dijkstra et al. (1989) investigated sublexical cross-modal activation effects by means of a two-choice forced-response task involving letters and auditorily presented syllables. The decision as to which vowel was presented in the syllable (e.g., /a/ or /e/ in the syllable /pa/) was facilitated with respect to a control condition when this vowel was primed by the more or less simultaneous presentation (ranging from 250 ms before or after vowel onset) of a congruent letter (e.g., A). In contrast, when the letter was congruent with the vowel associated with the other response (e.g., letter E with the syllable /pa/), the reaction times (RTs) were slower. These results were interpreted as evidence for fast and automatic activation effects from grapheme to phoneme representations. Because this study focused on visual-to-auditory effects only, further research concerning phoneme-to-grapheme effects is needed.

In other bimodal research, Frost and Katz (1989) presented subjects simultaneously with spoken and printed words or nonwords in English or Serbo-Croatian and asked for a same/different matching response. Within each language the effects of visual or auditory degradation were measured relative to an undegraded presentation. The visual stimuli were degraded by the overlay of a random dot pattern, and the auditory stimuli were masked with signal-correlated noise. The results for words and nonwords showed that degradation of either speech or print had a more deleterious effect on matching performance in the orthographically deep English than in the shallower Serbo-Croatian. The authors drew two conclusions from their findings. First, because comparable effects were obtained for high- and low-frequency words and nonwords, cross-modal activation was claimed to take place sublexically. Second, on the basis of the differences in compensatory effects for the degraded stimuli between Serbo-Croatian and English, the interaction between phonological and orthographic representations was assumed to be bidirectional. The smaller effect of both visual and auditory degradation in Serbo-Croatian was assumed to originate from the difference in the complexity of the connections between the orthographic and phonological representation systems (high complexity in English vs. low complexity in Serbo-Croatian).

It should be noted that this conclusion concerning the bidirectionality of cross-modal activation seems to be at variance with an assumption made by Frost and Katz (1989) in their introduction. There they proposed that subjects perform the matching task by recoding the orthographic information into a phonological representation and minimized the contribution of activation transfer in the opposite direction, which in fact must take place for the effects to be bidirectional. In our opinion, because it was unclear whether the subjects based their matching decision on an orthographic code, a phonological code, or even both (cf. Wood, 1974, 1977), it is impossible to draw conclusions about directionality. As a consequence, the relative amount of activation

flowing in the two directions could not be determined in this study.

Whereas the studies just reviewed explicitly focused on the issue of grapheme-to-phoneme effects between sublexical representations, a few other studies have examined cross-modal effects in the general context of word recognition, providing an indirect source of experimental evidence for phoneme-to-grapheme effects (Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981; Seidenberg & Tanenhaus, 1979; Tanenhaus, Flanigan, & Seidenberg, 1980). For example, Tanenhaus et al. (1980) observed that color naming latencies were longer (compared with control conditions) when the target words were preceded by auditory prime words that were phonologically similar. The temporal locus of these effects, however, was not determined in these studies.

To summarize, there are a limited number of studies indicating that automatic sublexical grapheme–phoneme interaction effects can occur. Most of these studies have focused on grapheme-to-phoneme interaction. The temporal aspects of the cross-modal activation process have not been investigated in much detail, and prelexical phoneme-to-grapheme effects have not been convincingly demonstrated. One reason for the lack of research on this last type of effect may be that, whereas the dependence of visual word recognition on auditory processing systems can be motivated by phylogenetic and ontogenetic arguments (Scinto, 1986), the opposite is harder to maintain (see Ehri, 1985, for an alternative view involving a mutual dependence).

Although the results of both the visual and bimodal studies have been collected by means of a variety of experimental tasks, they have all been interpreted within one particular model type: one that assumes language-specific links between orthographic and phonological sublexical representations. For example, the parallel distributed processing (PDP) model by Seidenberg and McClelland (1989) accounts for phonological effects in visual word recognition in terms of facilitatory connections between phoneme triples (Wickelfeatures; Wickelgren, 1969) and letter triples. In this model, no inhibitory connections are assumed to exist within or between the orthographic or the phonological code systems.

A similar connectionist model has been proposed by Lukatela and his colleagues (Lukatela, Carello, & Turvey, 1990; Lukatela, Turvey, Feldman, Carello, & Katz, 1989) to explain the results of a series of experiments on visual word recognition in different orthographies. In addition to facilitatory connections between graphemes and associated phonemes, this model assumes, like the interactive activation model by McClelland and Rumelhart (1981), inhibitory connections between grapheme representations both within and between the two writing systems (e.g., V in Roman and B in Cyrillic).

Neither of these two implemented models for visual word recognition incorporates phoneme-to-grapheme connections. Both therefore implicitly assume that activated phonemes do not directly affect graphemic processing. Frost and Katz (1989), however, described a network model, developed to explain the matching data presented earlier, that is not directionally constrained. However, because their model is not implemented, it is impossible to conduct simulations

to study phoneme-to-grapheme activation and its role in word recognition.

In conclusion, neither the available experimental evidence nor the existent word recognition models are very specific about the time course of activation of grapheme and phoneme representations and the directionality of their activation in visual or auditory word recognition. Clearly, there is more work to be done both on the experimental and the modeling fronts.

In this article, we present some experimental results that provide further insight into the nature of cross-modal activation. Exploiting experimental paradigms and data analysis techniques used thus far only to study the integration of non-linguistic stimuli, we investigated (a) whether (nonlexically mediated) facilitation effects can be obtained between graphemes and phonemes; (b) whether not only grapheme-to-phoneme activation, but also phoneme-to-grapheme activation, exists; and (c) whether cross-modal activation is facilitatory only (as the connectionist models just described assume) or can be inhibitory as well.

To conduct the experiments, we made use of a bimodal detection task in which subjects monitored two information sources in different modalities and gave a speeded detection response to a previously specified signal on either or both channels. Subjects detected visual targets such as letters (A and U) and/or auditory targets such as speech sounds (/a/ and /u/).[1] Both the structural and temporal relationships between the visual and auditory targets were varied in the experiments.

The structural manipulation involved comparing the subjects' latencies to detect targets in two different redundant conditions: congruent and incongruent conditions. In the former, congruent letters and speech sounds (e.g., letter A and sound /a/) were presented and in the latter incongruent letters and speech sounds (e.g., letter U and sound /a/). If there is cross-modal facilitation between graphemes and associated phonemes, faster RTs may be expected in congruent bimodal conditions, provided that these are equivalent with incongruent conditions in all respects except congruency.

To determine whether the observed differences between the congruent and incongruent conditions are due not just to cross-modal facilitation but also to inhibition, we introduced an additional redundant condition with a nonletter symbol (e.g., *) as the visual target stimulus. Because there is no phonological representation for this symbol and such a symbol fulfills different functions than does a letter or sound, we assumed that no cross-modal inhibition at a representation level was possible in this case.

We also manipulated the temporal relationship (stimulus onset asynchrony [SOA]) between the presentation of visual and auditory stimuli to examine the direction of the influence, that is, grapheme-to-phoneme and phoneme-to-grapheme activation. Assuming that single-channel RTs are similar and that a subject generally responds to the first presented target (cf. Colonius, 1990), RTs in conditions in which a visual target is followed by an auditory one should allow a test for phoneme-to-grapheme activation. Inversely, RTs to an auditory target followed by a visual one should reflect grapheme-to-phoneme activation. If the cross-modal influ-

ence of graphemes on phonemes is greater than that of phonemes to graphemes, more facilitation may be expected in the RTs when the first target is auditory than when it is visual.

Direct comparison of speeded bimodal detection responses in congruent and incongruent conditions at various SOAs is not possible, because congruent and incongruent bimodal conditions consist of different combinations of visual and auditory stimuli, each possibly different in terms of its processing characteristics. If the single-channel stimuli vary both within and between modalities as a result of characteristics such as frequency or saliency (Appelman & Mayzner, 1981), the amount of overlap between the RT distributions in bimodal conditions will vary as well. Only when the effects of stimulus differences across conditions are accounted for can congruent and incongruent conditions be sensibly compared.

Applying data analysis techniques used to study the integration of nonlinguistic stimuli makes it possible to correct the measured bimodal RTs for such stimulus differences. To explain how this can be done, we must refer to two classes of models that have been proposed to explain how subjects process nonlinguistic bimodal input in the bimodal detection task: separate activation models and coactivation models (Miller, 1982). Both types of models have been developed to explain the redundant signals effect (RSE), which refers to the finding that the average RTs to a bimodal stimulus are generally faster than those to either single-channel stimulus alone.

For the class of separate activation models, the RSE is considered to be the result of a race between two temporally overlapping detection processes of randomly varying durations in each channel. According to such models, inputs on different channels are processed separately but in parallel. Both channels compete in collecting stimulus evidence (called "activation" by Miller, 1982) and respond as soon as a target is detected in either. If the distributions of the detection responses in the two channels overlap so that each channel wins the race on some of the redundant trials, then the responses to the redundant signals will, on average, be faster than those to the single-channel conditions. The size of this effect of "statistical facilitation" (Raab, 1962) depends on the shape and overlap of the distributions of the processes involved.

In coactivation models, activation from different channels may be combined during processing. Because activation is assumed to build up gradually over time until the criterion is reached, two channels combining their activation will, on average, lead to responses faster than those to only one channel; in other words, an RSE will be observed.

Whereas separate activation models exclude coactivation as a contributing factor to the RSE, coactivation models do not exclude statistical facilitation (Miller, 1982, 1986). It is a common assumption in both types of models that processing a stimulus involves a number of steps, each of which may

---

[1] We use the notation /u/ rather than the International Phonetic Alphabet /y/ for the phoneme that is pronounced in French as *lune* or in German as *über*, to make it more congruent with the visual notation U.

take a variable amount of time. This variability leads to statistical facilitation in the RTs to redundant signals that are temporally close (cf. Ulrich & Giray, 1986).

Both separate activation models and coactivation models predict an RSE, but they differ in how large an effect they predict. Miller (1982, 1986) has developed an analysis technique that can specify the limits of the facilitation predicted by separate activation models. More facilitation than predicted argues for the existence of coactivation, and the entire class of separate activation models can be rejected. However, if less than the predicted facilitation is obtained, it is impossible on the basis of Miller's tests (described in the Appendix) to distinguish between separate activation and coactivation models.

In this article, we shall show that the predictions of both separate activation models and Miller's (1982, 1986) coactivation test can fruitfully be used to test psycholinguistically motivated coactivation models that assume cross-modal contacts at a representation level (in fact, all of the psycholinguistic models reviewed earlier belong to this class). First we indicate how the predictions of a separate activation model can be used to correct RTs for differences between stimuli. Then we show how Miller's coactivation test can be applied to support the existence of cross-modal effects at a representation level.

To correct for differences in statistical facilitation effects across conditions, the obtained RTs are adjusted through application of the prediction method underlying separate activation models. This involves subtracting from the obtained bimodal RTs the RTs predicted on the basis of an independent separate activation model. Once the bimodal RTs are corrected for differences in stimulus processing characteristics, the congruent conditions can be compared with the incongruent ones. Faster adjusted RTs in the congruent conditions than in the incongruent conditions indicate cross-model activation. In other words, compared with the single-channel conditions, a reduced RSE is expected for the incongruent conditions, because no grapheme–phoneme links exist that can cause coactivation at the representation level.

Miller's (1982, 1986) test for coactivation can be applied to the bimodal response distributions to provide an additional test of the hypothesis of cross-modal activation. If coactivation is indeed present, the RTs obtained in the congruent conditions should in general be faster than those predicted under conditions of maximal separate activation and thus not be compatible with any race model. Because of the absence of coactivation effects at the representation level, fewer violations of the predicted distribution curves are expected in the incongruent conditions, although coactivation at decision level or response level could still occur.[2]

We applied the two techniques just described in three experiments to study the bimodal integration of graphemes and phonemes. In Experiment 1, we attempted to collect some initial evidence in favor of coactivation at a representation level by presenting visual and auditory targets at an SOA of 0 ms. We assumed that a simultaneous presentation of visual and auditory signals would lead to the largest overlap of processing and thus, potentially, to the largest coactivation effects. In Experiment 1, we also tested the hypothesis that

phoneme representations can activate those of graphemes. To do so, we included bimodal stimuli in which the visual target preceded the auditory one by 100 ms (denotated as SOA = −100 ms).

In Experiment 2, we used a different mode of stimulus presentation (mixed instead of blocked) and incorporated one extra SOA (SOA = 100 ms) to allow the examination of grapheme-to-phoneme activation effects. This mixed design was used again in Experiment 3 to obtain additional support for bidirectional effects. The results of all three experiments were analyzed by means of the techniques just mentioned and described in more detail in the Appendix.

Experiment 1

In the present experiments, subjects were asked to respond as quickly as possible if they detected a previously specified letter or symbol (e.g., A, U, or *) and/or certain speech sound (e.g., /a/). The obtained bimodal RTs were first corrected for statistical facilitation effects by subtracting the RTs predicted on the basis of independent activation of the stimuli in each of the channels separately. Next, the adjusted RTs to congruent bimodal stimuli (e.g., letter A and sound /a/ in Dutch, abbreviated henceforth as Aa) were compared with those to incongruent stimuli (e.g., letter U and /a/, abbreviated Ua) and those to bimodal stimuli where one target was not a letter but a symbol (e.g., * and /a/, abbreviated *a).

In Experiment 1, bimodal stimuli in all conditions were presented at two SOAs in order to investigate the occurrence of phoneme-to-grapheme activation and the presence of coactivation effects. At an SOA of 0 ms, the visual and auditory stimuli had the same onset; at an SOA of −100 ms, the onset of the visual stimulus preceded that of the auditory one by 100 ms.[3]

The largest coactivation effects in bimodal trials are expected to arise at SOAs that compensate for differences in processing times between single channels (Colonius, 1990; Miller, 1986). Because different visual and auditory stimuli probably vary in several respects (frequency, saliency, modality-dependent processing, etc.), differences in processing time are to be expected. However, the available experimental literature does not provide any detailed information concerning the processing times of the various target letters

---

[2] Some problems with the interpretation of data obtained with the bimodal detection task have been noted, particularly in terms of fast guessing (Eriksen, 1988) and response preference strategies (Mullin, Egeth, & Mordkoff, 1988). However, it seems reasonable to assume that these general strategies should influence the experimental results similarly for congruent and incongruent conditions and thus be of little consequence for our interpretation.

[3] A negative SOA indicates a condition where a visual stimulus precedes an auditory stimulus; a positive SOA indicates a condition in which an auditory stimulus leads. Reaction time, however, is always measured from the onset of the first presented target (either visual or auditory). For example, an SOA of −100 ms for Condition Ia stands for a condition in which visual stimulus I precedes auditory stimulus /a/ by 100 ms. If I is not a target, but /a/ is, RT is measured from the onset of /a/.

and phonemes. For stimuli whose processing distributions are similar in shape and temporal characteristics, the largest coactivation effects are expected at an SOA of 0 ms. An SOA of 0 ms thus seems to be a good first choice.

To study phoneme-to-grapheme effects, we presented visual targets 100 ms before the auditory ones in bimodal trials. The choice of this SOA was motivated by two considerations. On the one hand, the longer the (negative) SOA, the more likely the experiment would provide evidence for cross-modal influences that originate from earlier phases of auditory processing and are thus more likely representational in nature. On the other hand, the SOA should not be too large, in order to allow cross-modal effects to occur at all.

Method

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

Design. The experiment consisted of three sets of experimental trials, each of which involved a different combination of instruction and stimulus material (hereafter referred to as a target set). Target sets included a visual stimulus only, an auditory stimulus only, or both a visual and an auditory stimulus. In all target sets the letter E and sound /e/ occurred as nontarget stimuli, whereas the sound /a/ was always the auditory target. Target sets differed with respect to the visual target stimuli in the single-channel and redundant (bimodal) conditions. In Target Set Aa, the visual target stimulus was the letter A; in Target Set Ua, the letter U; and in Target Set *a, the symbol *. No trials combined target with nontarget stimuli. Table 1 summarizes the combinations of visual and auditory stimuli in the various go and no-go conditions.

Furthermore, in each of the target sets, the redundant trials were presented at two SOAs. For an SOA of –100 ms, the visual stimulus appeared 500 ms after the warning signal and 100 ms before the auditory stimulus was started. For an SOA of 0 ms, the visual and auditory stimuli had simultaneous onsets, 600 ms after the warning signal. In the single-channel visual condition, the visual stimulus was presented 500 or 600 ms after the warning signal, in correspondence with the two SOAs of the redundant conditions. The auditory stimulus always appeared 600 ms after the warning signal. In order to balance the number of visual and auditory trials, equal numbers of single-channel auditory trials were allocated to the –100-ms and 0-ms SOA conditions.

Table 1
Stimulus Conditions in Experiment 1

Target set	Condition		
	Visual	Auditory	Redundant
Aa			
Go	A–	–a	Aa
No go	E–	–e	Ee
Ua			
Go	U–	–a	Ua
No go	E–	–e	Ee
*a			
Go	*–	–a	*a
No go	E–	–e	Ee

Note. The first letter of a condition refers to the visual stimulus, the second to the auditory; a dash indicates no signal.

Single-channel (auditory only and visual only) trials and bimodal trials were repeated 40 times. Each of the three target sets (Aa, Ua, and *a) was presented in a separate session and had the following dimensions: 3 (type of trial) × 2 (target/nontarget stimulus) × 2 (SOA) × 40 (repetitions) = 480 test stimuli. An additional 48 practice trials were constructed, leading to a total number of 528 trials for each target set.

Stimuli. The auditory stimuli were recorded on tape by a female native speaker of Dutch in a soundproof room. The duration of the vowel /a/ was 280 ms, whereas that of /e/ was 350 ms. The stimuli were digitized on a VAX 11/750 computer with a sampling rate of 20 kHz. For each target set, a randomized sequence of targets and nontargets 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, inaudible, channel of the tape, a pulse was placed that triggered both the timer for the recording of the RTs and the presentation of the visual stimulus.

The visual stimuli were white Roman capital letters, 6 mm in height, presented on a MATROX screen with a dark background. To provide projection within the foveal field of the eye, we placed the monitor at a distance of 60 cm from the subject, resulting in a visual angle of approximately half a degree. All visual stimuli were presented for 280 ms, the duration of the auditory /a/ target stimulus. Presentation of the visual stimuli and recording of the RTs were controlled by a PDP-11/23 computer. The asterisk was chosen as a neutral stimulus since it has no obvious relation to the phonemes.

Procedure. Subjects participated in three sessions, each consisting of one target set, on successive days. The order of target sets in these sessions was counterbalanced over subjects. Before the presentation of a target set, subjects read written instructions. The instructions were repeated orally at the beginning of the experiment. Subjects were told to rest the index finger of their preferred hand lightly on the response button in front of them and to push this button as fast as possible whenever they saw and/or heard a target stimulus. At the beginning of the presentation of Target Set Aa, they were told to react when they saw the letter A, when they heard the sound /a/, or when both the letter A and the sound /a/ appeared. Before the presentation of Target Set Ua, they were instructed to react whenever they saw the letter U and/or heard the sound /a/; before Target Set *a, when the symbol * and/or the sound /a/ occurred. Each time they were also told not to respond to presentations of letter or speech sound E. The task was therefore a go/no-go one.

Each trial started with a 1000-Hz warning signal 200 ms in duration. In the single-channel visual condition, this warning signal was followed by 500 or 600 ms of silence before the visual signal was presented. In the single-channel auditory condition, the period of silence between warning signal and auditory stimulus always was 600 ms. The redundant trials combined the two presentations, resulting in an onset of the visual signal 100 ms before or simultaneous with that of the auditory signal. Two seconds after presentation of the last signal, a new trial was initiated.

Each experimental session lasted about 45 min, with a short break after about 25 min. Each session consisted of 48 practice trials followed by a block of 480 test trials. After the practice trials, there was a short pause in which subjects could ask for clarifications if necessary.

Results

Mean RTs (measured from the onset of the first presented target stimulus) were computed for each subject and each experimental condition in each target set. Latencies longer than 750 ms or shorter than 150 ms were treated as errors.

The total percentage of missing and extreme values was 1.1%. Errors were substituted by mean RTs in each subcondition for each subject. The percentage of false alarms, that is, reactions to the no-go trials, was 2.1%. Table 2 shows the main results for each SOA separately.

To determine whether the RSE was significantly larger for an SOA of −100 ms than for an SOA of 0 ms, we performed an analysis of variance (ANOVA) on the RT differences between the redundant conditions and the visual single-channel condition for all three target sets and for both SOAs. This ANOVA with the factors SOA and target set showed a significant main effect of SOA, $F(1, 30) = 117.52, p < .001$, but no main effect of target set, $F(2, 60) < 1, p > .50$, and no interaction between SOA and target set, $F(2, 60) = 1.24, p > .10$.

As explained in the Introduction, performing an ANOVA on the raw RT data would not have taken into account the processing differences between the various target letters and speech sounds. Instead, we first compared the RTs obtained for the different target sets with those predicted by an independent race model and then tested for the presence of coactivation effects in the data by means of the test proposed by Miller (1982).

Independent race predictions. If coactivation existed in Target Set Aa, which at least in part can be located at a representation level, a larger RT deviation could be expected to arise from the predictions of an independent race model for Target Set Aa than from the predictions of such a model for the other target sets. The comparison of the redundant conditions in the different target sets must take differences into account resulting from statistical facilitation. Thus, for each subject and each redundant condition we first computed the predicted minimum distribution using Inequality A1 in

the Appendix. After adding 100 ms to the RTs of the single-channel auditory condition at an SOA of −100 ms, we ordered the correct RTs in millisecond steps for the analysis to obtain the highest resolution in the predictions. Only RTs between 150 and 750 ms were included in the analysis. The mean predicted bimodal RTs computed from the predicted minimum distribution are given in Table 3 for the SOAs of −100 ms and 0 ms, together with the means obtained, the significance of their difference, and the correlation between the means obtained and predicted.

For each subject, the mean obtained RT for a certain redundant condition under a specific SOA was subtracted from each RT predicted for that condition and with that SOA. After this correction for statistical facilitation, the resulting times were used as estimates of the amount of representational coactivation. To test for differences in the amount of coactivation among target sets, we performed an ANOVA on the adapted redundant conditions. This analysis showed a main effect of target set, $F(2, 60) = 8.27, p < .001$, but not of SOA and no interaction between target set and SOA (both Fs < 1). Thus, the SOA = − 100 ms and SOA = 0 ms conditions did not differ significantly in the size of facilitation effects with respect to an independent separate activation model.

Disregarding SOA, paired planned comparisons were performed on the differences between the adapted redundant conditions, referred to as Aa', Ua', and *a'. These comparisons showed significant differences between Target Sets Aa' and Ua', $t(30) = -4.14, p < .001$, and between Aa' and *a', $t(30) = -2.17, p < .05$. The difference between Ua' and *a' was marginally significant, $t(30) = 1.83, p = .08$.

Coactivation test. Next, we tested whether coactivation effects could be demonstrated by applying Miller's (1982, 1986) test to all conditions. However, before applying this test to the SOA = 0 ms condition (simultaneous presentation), we followed Miller's (1982) suggestion to compute the average across subjects for the faster of the two single-channel conditions in a target set. The three resulting values were 328 (Target Set Aa), 341 (Target Set Ua), and 343 (Target Set *a). Testing against the obtained values in the redundant conditions of 290, 317, and 310, respectively, we found significant differences in all cases: For Target Set Aa, $t(30) = 8.46, p < .001$; for Target Set Ua, $t(30) = 6.43, p < .001$; and for Target Set *a, $t(30) = 9.61, p < .001$. This result indicates that RSEs obtained in each target set were not an artifact of averaging across some subjects who detected the visual signal faster and other subjects who detected the auditory signal faster (Miller, 1982). (Such a detection difference would be quite unlikely at an SOA of −100 ms, where the visual signal appeared 100 ms before the auditory signal; therefore no such test was performed for this SOA.)

To check for coactivation effects, we computed the average cumulative distribution functions (CDFs) for the single-channel visual and auditory conditions and the redundant conditions in each target set and at each SOA by averaging across subjects (Ratcliff, 1979). To obtain these CDFs (depicted for an SOA of 0 ms in Figures 1–3), we rank ordered the RTs in each condition for a given subject (for the SOA of −100 ms, after adding 100 ms to the RTs in the single-channel auditory condition). If any of the 40 RTs were below

Table 2

*Mean Reaction Times (RTs; in Milliseconds) for the Redundant and Single-Channel Conditions in Target Sets Aa, Ua, and *a at Two Stimulus Onset Asynchronies (SOAs)*

Target set	Condition		
	Visual	Auditory	Redundant
SOA = −100 (letter led by 100 ms)			
Aa	345	345	328
Ua	376	348	360
*a	368	351	351
SOA = 0 (simultaneous onset of letter and speech sound)			
Aa	340	345	290
Ua	373	351	317
*a	358	360	310

Note. RTs were measured from the first target stimulus. RTs to single-channel auditory trials were to identical tokens presented 600 ms after the ending of the warning signal, RTs to single-channel visual trials were to identical tokens presented at 500 ms (SOA = −100 ms) or 600 ms (SOA = 0 ms) after the warning signal ended.

Table 3

Mean Reaction Times (in Milliseconds) Predicted by an Independent Separate Activation Model and Obtained for Redundant Conditions in Target Sets Aa, Ua, and *a at Two Stimulus Onset Asynchronies (SOAs) and t Tests and Correlations for the Predicted and Obtained Means

Target set	Predicted	Obtained	Difference	$t(30)$	p	r
		SOA =	−100 (letter led by 100 ms)			
Aa	335	328	7	−2.57	<.05	.96
Ua	358	360	−2	.60	ns	.94
*a	354	351	3	−1.25	ns	.96
		SOA = 0 (simultaneous onset of letter and speech sound)				
Aa	301	290	11	−2.82	<.01	.92
Ua	315	317	−2	.35	ns	.93
*a	312	310	2	−.83	ns	.93

Note. All correlations were significant at $p < .001$.

150 ms or above 750 ms, a full distribution was generated by means of a damped cubic SPLINE function (cf. De Boor, 1978). Each of the 40 ordered RTs estimates the RTs at the 1.25th, 3.75th, and 6.25th to 98.75th percentiles of the true CDF for a given subject. Composite CDFs were then formed by averaging, across subjects, all the RTs for a given percentile (Miller, 1982). The redundant-signal CDFs were compared with the sum of the single-channel CDFs, represented in Inequalities A3 and A4 in the Appendix. For an SOA of −100 ms, Inequality A4 was violated throughout the range from the 1.25th to the 61.25th percentiles of RT for Target Set Aa, with the exception of the 3.75th percentile,

$t(30) = .19$, and from the 6.25th to the 46.25th percentiles of RT for Target Set *a, as demonstrated by paired t tests between redundant-signal and sum-curve distributions across subjects at each of the 40 percentile points in a session. Inequality A4 was never significantly violated under this SOA for Target Set Ua. At an SOA of 0 ms, Inequality A3 was violated from the 1.25th to the 61.25th percentile for Target Set Aa, from the 1.25th to the 26.25th percentile for Target Set Ua, and from the 7.75th to the 33.75th percentile for Target Set *a. Compared with the size of the violations reported in the literature (Miller, 1982, 1986), the violations in Target Set Aa are very large. Our tests therefore indicated
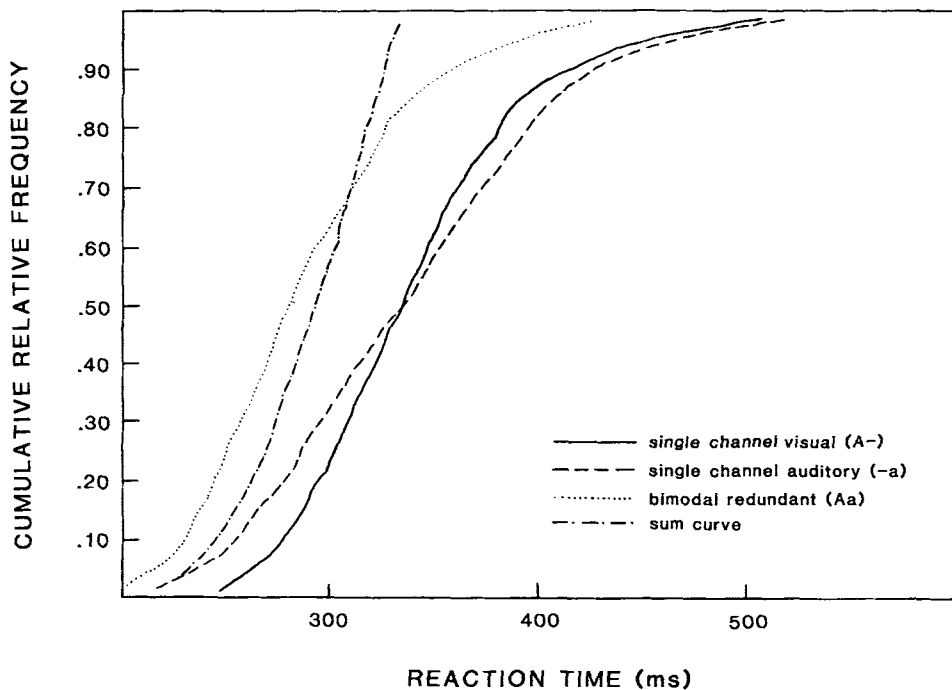

Figure 1. Cumulative distribution functions (CDFs) for redundant, single-channel visual, and single-channel auditory conditions, and the sum of the single-channel CDFs, for Target Set Aa at a stimulus onset asynchrony of 0 ms.
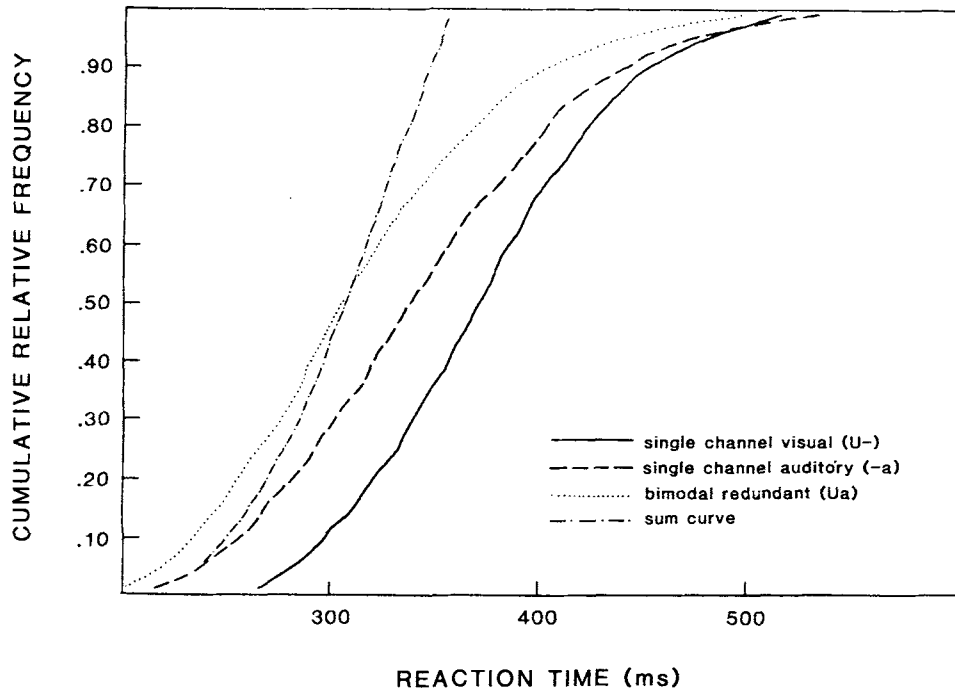
Figure 2. Cumulative distribution functions (CDFs) for redundant, single-channel visual, and single-channel auditory conditions, and the sum of the single-channel CDFs, for Target Set Ua at a stimulus onset asynchrony of 0 ms.

Figure 3. Cumulative distribution functions (CDFs) for redundant, single-channel visual, and single-channel auditory conditions, and the sum of the single-channel CDFs, for Target Set *a at a stimulus onset asynchrony of 0 ms.
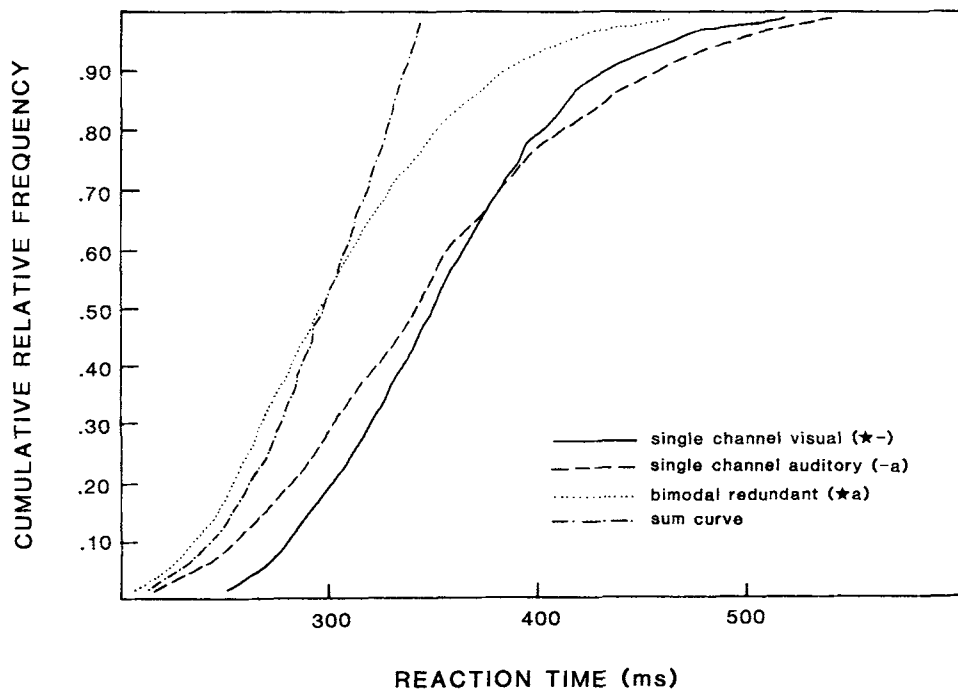
that coactivation was present in all three types of target sets, and most strongly in Target Set Aa.

Finally, ANOVAs were performed on the raw data to test for differences among the single-channel conditions across SOAs and target sets. For the single-channel auditory conditions, no significant main effect of target set was found, $F(2, 60) < 1, p > .50$, indicating that these single-channel conditions were comparable over target sets. Nor was a significant difference found between single-channel auditory trials at SOAs of -100 ms and 0 ms, $F(1, 30) = 2.73, p > .10$, as was to be expected, because all trials consisted of identical tokens. The interaction between the target set and SOA was not significant either, $F(2, 60) = 2.40, p = .10$. For the single-channel visual conditions, different results were obtained. Significant differences were found among target sets, $F(2, 60) = 13.79, p < .001$, and between the SOA $= -100$ ms and SOA $= 0$ ms conditions, $F(1, 30) = 12.80, p < .001$, for which the moment of stimulus onset (at 500 or 600 ms after the warning signal) differed by 100 ms. The interaction between target set and SOA was not significant, $F(2, 60) = 1.35, p > .10$.

Discussion

The results of Experiment 1 confirm the existence of cross-modal interactions between phoneme and grapheme representations for SOAs of -100 ms and 0 ms. For both SOAs, the differences between obtained and predicted RTs were significantly larger in Target Set Aa than in the other sets. In addition, application of Miller's (1982, 1986) test to this target set for both SOAs also showed a violation of the sum-curve distribution over a much longer range than for Target Sets Ua and *a.

Our finding of cross-modal activation effects under an SOA of -100 ms supports the claim that we have measured fast representational effects. We may illustrate this by the following rough calculation. Subtracting the SOA from the bimodal RT, and assuming that motor processes occurring after perceptual and decision processes take on the order of 90 ms or longer (cf. Luce, 1986; MacKay & Bonnet, 1990), the visual and auditory representations already contact each other before about $328 - 100 - 90 = 138$ ms of auditory processing.

The results not only indicate the presence of coactivation at a representation level, but, more specifically, suggest the activation of graphemes by congruent phonemes. In contrast to the different single-channel RTs obtained by Miller (1986) for plus signs and tones, our single-channel RTs for letters and auditory vowels indicate that letters and vowels were processed with equal speed. Given this finding of approximately equal single-channel RTs, it seems plausible (as was argued at the end of the Introduction) that subjects in the redundant conditions reacted predominantly to the letter, which preceded the auditory signal by 100 ms. Because larger facilitation effects were found in the corrected Target Set Aa than in the other sets, the graphemic representations on which the responses were based in this target set must have been influenced by their phonemic counterparts.[4]

It is important to note that this cross-modal activation must have taken place during the identification of the visual stimulus and not when the decision to respond was made. Explanations localizing the coactivation effects solely at the decision or response levels can be excluded, since the task demands in Target Sets Aa and Ua were comparable. Indeed, the visual targets in both target sets (letter A and letter U) were associated with the go response.[5]

The conclusion that graphemes can be activated quickly by congruent phonemes is of considerable interest, because this direction of sublexical cross-modal effects has not been investigated in much depth. Although these cross-modal activation effects are not directly predicted by the word recognition models reviewed in the Introduction (e.g., Frost & Katz, 1989; Lukatela et al., 1990; Lukatela et al., 1989; Seidenberg & McClelland, 1989), they can be easily accounted for by expanding the models to allow for phoneme-to-grapheme activation.

Experiment 1 produced another result that is directly relevant to these models. The RT patterns in the Ua target set did not differ from those in the *a target set. This suggests a lack of inhibition between phonemes and incongruent graphemes (or the phonemes activated by those graphemes) in Target Set Ua, under the assumption that a nonletter symbol like '*' is not inhibited by any phonemes.

Comparing the patterns of the raw RTs for SOAs of -100 ms and 0 ms, we observed a strong increase in the RSE from the -100-ms SOA to the 0-ms SOA in all target sets. The close predictions of the RTs by means of an independent race model for both SOAs indicate that a large part of this increase must have been due to statistical fa-

---

[4] It is possible to construe accounts that explain facilitatory effects in congruent conditions in terms of preactivation of the later arriving targets (grapheme or phoneme) by the earlier targets (phoneme or grapheme). However, such accounts are inferior to ours for several reasons. First, Dijkstra (1990, Experiment 7) showed that subjects making a modality decision (i.e., whether the first stimulus presented was visual or auditory) gave 70% of their reactions to the visual channel under an SOA of -100 ms. (The stimuli were identical to the ones used in the present experiments.) In the present experiments, there is evidence of a similar distribution. The single-channel CDFs indirectly reflect the (expected) distribution of visual and auditory reactions to bimodal trials. Taking into account the SOA of -100 ms, the resulting CDFs indicate that for the RT value below which 70% of the visual reactions have been given, only 30% of the auditory reactions have arrived. Therefore, it is likely that a 70:30 distribution of reactions exists to visual and auditory stimuli, respectively, in bimodal trials at an SOA of -100 ms. However, whatever mechanism results in the observed coactivation, bidirectionality is a necessary conclusion because significant facilitation occurred for both an SOA of -100 ms and an SOA of 100 ms, as Experiments 2 and 3 show.

[5] The same line of reasoning was followed by Miller (1991). As he remarked, "Response interference cannot be the source of the effects . . . , because these effects reflect differential processing in various conditions in which response processing is equated (i.e., both channels contain targets supporting the same response)" (pp. 167–168).

cilitation. The increase was to be expected on the basis of the larger overlap of visual and auditory distributions under 0-ms SOA. The importance of statistical facilitation is further strengthened by replicated finding of high correlations between the obtained and predicted RTs for the redundant conditions (all above .90).

Apart from statistical facilitation, coactivation at a representation level must underlie the RSE in the congruent redundant conditions for both the −100-ms and the 0-ms SOAs. First, the differences between the corrected congruent and other redundant conditions were statistically significant, as were the differences between the obtained RTs and those predicted by an independent race model (of similar magnitude under both SOAs). Furthermore, a consistent large violation of the sum-curve was found for the congruent conditions under both SOAs.

A comparison of the size of the cross-modal activation effects under the two SOAs is difficult to make, even on the basis of the corrected RTs. The responses under 0-ms SOA were an unknown mixture of reactions to the visual and to the auditory target, whereas those under −100-ms SOA (as argued before) probably consisted mainly of reactions to the visual target. It is thus conceivable that larger (phoneme-to-grapheme) coactivation effects in reactions to the visual target at the 0-ms SOA were averaged out by the presence of smaller (grapheme-to-phoneme) coactivation effects in auditory target reactions.

The important conclusion remains that the results of Experiment 1 were in favor of the hypothesis of cross-modal activation. Of course, it is important to exclude alternative explanations. For example, one type of explanation for the pattern of results could be construed in terms of differences in the visual discriminability of the various target and nontarget stimuli: Perhaps the visual target letter A appeared less similar to the nontarget E than did the target U, leading to faster RTs for single-channel visual and redundant conditions in the Aa target set than in the Ua target set. However, because the prediction method takes into account differences in RTs between the single-channel conditions, this perceptual bias hypothesis would still not explain why the facilitation effect with respect to the race prediction would be larger in the Aa target set than in the other sets.

Furthermore, as a test for the existence of such differences in target discriminability, we performed a control experiment that replicated Experiment 1, except that only visual target and nontarget stimuli were included. No RT differences were obtained between blocks of trials with A and U target letters, both mixed with the nontarget letter E. This strongly suggests that the two target letters did not differ in visual discriminability from the nontargets.

We feel that differences in visual discriminability among targets and nontargets cannot explain the pattern of results, even though the RTs to the single-channel stimuli differed significantly between the congruent (Aa) and incongruent (Ua) target sets (as can be seen in Table 2). Nonetheless, to exclude any potential problems caused by the blocked presentation, we performed a second experiment, introducing a mixed presentation of stimuli. A mixed design would allow

us to predict both congruent and incongruent bimodal RTs on the basis of the same single-channel RTs. If the effects found in Experiments 1 were genuine, automatic representational effects, they would remain robust in different experimental designs.

However, our most important motivation in conducting a second experiment was not technical but theoretical in nature. Earlier, we argued that reactions to a visual target followed by a congruent auditory target 100 ms later predominantly reflect auditory-to-visual activation. Similarly, an auditory target followed by a 100-ms delayed congruent visual target should reflect visual-to-auditory activation. A comparison of the results of these two SOA manipulations could thus potentially indicate a directional asymmetry in the size or speed of cross-modal activation (e.g., the influence from auditory to visual might be less strong than from visual to auditory). Such evidence with respect to the mutual dependence of visual and auditory sublexical processing systems should be useful in the construction of models simulating both visual and auditory word recognition processes (e.g., when a phonological input channel is added to the model for visual word recognition developed by Seidenberg & McClelland, 1989).

Thus, Experiment 2 included mixed SOA conditions in which the visual stimulus preceded the auditory one by 100 ms, was presented simultaneously with it, or followed it by 100 ms. This design made the experiment temporally symmetric with respect to the visual and auditory modality. We wanted to investigate whether this temporally more balanced design would lead to RT differences in the SOAs used before, as a result of differences in the subjects' division of attention over the two modalities.

Finally, in Experiment 2 a bimodal condition was added in which a target or nontarget in one modality was paired with a neutral stimulus in the other. A neutral stimulus was not a target stimulus and occurred equally often in go and no-go trials. The neutral stimuli used were the letter I, the speech sound /i/, a star, or white noise. The inclusion of this bimodal baseline condition served two important goals. First, a comparison of data from this condition with the raw data of the congruent and incongruent redundant conditions can be used to evaluate general effects of neutral linguistic and nonlinguistic stimuli on target stimuli in the other modality (Miller, 1982, Experiment 3). Second, RT differences between linguistic and nonlinguistic neutral stimuli may be informative with respect to the depth of processing of the nontarget stimulus and the presence of attention shifts from one modality to the other under different SOAs.

To summarize, Experiment 2 included bimodal redundant stimuli (e.g., visual A combined with auditory /a/, abbreviated Aa), bimodal nonredundant stimuli (e.g., visual I combined with auditory /a/, abbreviated Ia), and single-channel stimuli (e.g., visual A alone, abbreviated A-) presented under three different temporal relationships (SOAs of −100, 0, and 100 ms). Also, all conditions were included in one completely mixed experiment, in order to reduce variances due to session and learning effects. Finally, as exemplars of congruent redundant conditions both Aa and Uu trials were

included; Au and Ua trials made up the incongruent redundant conditions.

Experiment 2

Method

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

Design. The experiment was conducted in three sessions. In all sessions, both the go and no-go trials fell into one of 16 different stimulus presentation conditions, as shown in Table 4. There were four bimodal redundant conditions, eight bimodal nonredundant conditions, and four single-channel conditions for both go and no-go trials. In the redundant and nonredundant conditions three SOAs were used: The visual stimulus was presented 100 ms before (−100 ms), at the same time as (0 ms), or 100 ms after (100 ms) the auditory stimulus.

Redundant conditions were repeated 20 times under all SOAs. Single-channel conditions were repeated 20 times. This led to a total of 1,600 test stimuli: 12 (conditions) × 2 (target/nontarget stimulus) × 3 (SOA) × 20 (repetitions) + 4 (single-channel conditions) × 2 (target/nontarget stimulus) × 20 (repetitions) = 1,440 + 160 = 1,600 stimuli. Furthermore, 40 practice trials for each session were constructed so that the total number of stimuli presented in the experiment as a whole amounted to 1,720. The number of trials in a session was therefore 573 (or 574).

As before, the auditory stimuli were recorded on tape by a female native speaker of Dutch in a soundproof room. Naturally sounding stimuli of about equal length (320 ms) were chosen for use in the experiment. The noise stimulus consisted of a computer-generated and recorded white-noise signal of 320 ms, the average duration of the auditory stimuli. All visual stimuli were also displayed for 320 ms.

Procedure. The experiment was conducted over 3 successive days. The order of sessions was counterbalanced over subjects.

In this experiment, subjects were instructed to push the response button as fast as possible whenever they saw and/or heard the letter or sound "A" or the letter or sound "U" (in Dutch). They were told not to react if other letters or sounds were presented alone or in combination.

Each trial started with a 1000-Hz warning signal 200 ms in duration. In the single-channel visual and auditory trials, this warning signal was followed after 600 ms by the target stimulus. In the redundant trials, the visual stimulus followed the warning signal after 500, 600, or 700 ms of silence (depending on SOA); the au-

Table 4
Stimulus Conditions in Experiment 2

Visual	Auditory	Redundant		Nonredundant			
			Go				
A–	–a	Aa	Au	Ai	An	Ui	Un
U–	–u	Ua	Uu	Ia	Iu	*a	*u
			No go				
O–	–o	Oo	Oe	Oi	On	Ei	En
E–	–e	Eo	Ee	Io	Ie	*o	*e

Note. The first letter of a condition refers to the visual stimulus, the second to the auditory; n indicates white noise; a dash indicates no signal.

Table 5
Mean Reaction Times (in Milliseconds) for the Redundant and Single-Channel Conditions When the Visual Stimulus Preceded the Auditory Stimulus (Stimulus Onset Asynchrony [SOA] = −100 ms), Accompanied it (SOA = 0 ms), or Followed It (SOA = 100 ms)

Condition	Single-channel visual	SOA			Single-channel auditory
		−100 ms	0 ms	100 ms	
Aa	406	375	334	370	396
Au	406	408	379	403	443
Ua	422	410	364	409	396
Uu	422	402	369	399	443

Note. Reaction times were measured from the first target stimulus.

ditory stimulus was always presented after 600 ms. Two seconds after presentation of the auditory signal, a new trial was initiated.

Each session consisted of 40 practice trials followed by a block of 533 (or 534) test trials. After the practice set there was a short pause in which the subjects had an opportunity to ask questions. Each session lasted for about 45 min, with a 3-min break after about 25 min.

Results

Mean RTs (measured from the onset of the first presented target stimulus) were computed for each subject and experimental condition. Latencies longer than 750 ms or shorter than 150 ms were treated as errors. The total percentage of missing and extreme values was 1.3%. Errors were substituted by mean RTs in each subcondition for each subject. The percentage of false alarms (i.e., reactions to the no-go trials) was 1.6%. Table 5 shows the main results for the redundant and single-channel conditions.

Independent race predictions. We first present the analysis of the results for the bimodal redundant conditions; separate analyses are then presented for the nonredundant conditions. As in Experiment 1, we set out to compare the congruent and incongruent redundant conditions in the following way, in order to take differences resulting from statistical facilitation into account. For each subject and each condition, the predicted minimum distribution was computed, using the single-channel conditions according to Equation A1 in the Appendix.

After addition of 100 ms to the single-channel RTs of the later signal for the −100-ms and 100-ms SOA conditions, the correct RTs were ordered in millisecond steps for the analysis to obtain the highest resolution in the predictions. Only RTs between 150 and 750 ms were included in the analysis. The mean predicted bimodal RTs computed from the predicted minimum distributions given Equation A1 in the Appendix are shown in Table 6 for all SOAs, together with the means obtained, the size of their difference and its significance, and the correlation between the obtained and predicted means.

For each subject, the mean predicted RT for a certain redundant condition under a specific SOA was subtracted from each RT obtained in that condition and with that SOA. Table 6 indicates the differences between the predicted and ob-

Table 6

Mean Reaction Times (in Milliseconds) Predicted by an Independent Separate Activation Model and Obtained for Redundant Conditions Aa, Au, Ua, and Uu at Three Stimulus Onset Asynchronies (SOAs) and t Tests and Correlations for the Predicted and Obtained Means

Condition	Predicted	Obtained	Difference	$t(30)$	p	r
		SOA = −100 ms (letter led by 100 ms)				
Aa	394	375	19	−5.80	<.001	.90
Au	399	408	−9	2.51	<.05	.93
Ua	406	410	−4	.85	ns	.85
Uu	413	402	11	−2.90	<.01	.85
		SOA = 0 ms (simultaneous onset of letter and speech sound)				
Aa	358	334	24	−8.13	<.001	.92
Au	377	379	−2	.22	ns	.92
Ua	364	364	0	−.15	ns	.85
Uu	387	369	18	−5.16	<.001	.90
		SOA = 100 ms (speech sound led by 100 ms)				
Aa	385	370	15	−3.50	<.001	.89
Ua	386	409	−23	4.45	<.001	.87
Au	420	403	17	−4.74	<.001	.93
Uu	424	399	25	−7.39	<.001	.92

Note. All correlations were significant at $p < .001$.

tained RTs over SOA for the four redundant conditions. An ANOVA conducted using the adapted redundant conditions (later referred to as Aa′, Au′, Ua′, and Uu′) showed significant main effects of condition, $F(3, 90) = 23.81, p < .001$, and SOA, $F(2, 60) = 3.84, p < .05$, as well as a significant Condition × SOA interaction, $F(6, 180) = 19.20, p < .001$.

For the adapted data, we next wanted to test the various congruent and incongruent redundant conditions against each other for each SOA. Because the number of such comparisons in Experiment 2 (six per SOA) was much higher than in Experiment 1 (three per SOA), we decided to perform Newman-Keuls analyses with an alpha of .05, instead of planned paired comparisons. For SOAs of −100 ms and 0 ms, the following comparisons on the adapted conditions were significant: Aa′ versus Ua′; Aa′ versus Au′; Uu′ versus Au′, and Uu′ versus Ua′. For the SOA of 100 ms, the comparisons Aa′ versus Ua′, and Uu′ versus Ua′ were again significant, as was Au′ versus Ua′.

Coactivation test. We next tested for coactivation. Before applying Miller's (1982, 1986) technique to the SOA = 0 ms condition, we computed the average across subjects for the faster of the two single-channel conditions in a session. The four resulting values were 386 (Aa), 402 (Au), 389 (Ua), and 416 (Uu). Testing against the obtained values in the redundant conditions of 334, 379, 364, and 369, respectively, we found significant differences in all cases: For Condition Aa, $t(30) = 13.98, p < .001$; for Condition Au, $t(30) = 5.52, p < .001$; for Condition Ua, $t(30) = 4.94, p < .001$; and for Condition Uu, $t(30) = 11.10, p < .001$. This indicates that the RSE obtained in each session was not an artifact of averaging across some subjects who detected the visual signal faster and other subjects who detected the auditory signal faster (cf. Experiment 1).

The average CDFs for the single-channel visual and auditory conditions and the redundant conditions in each ses-

sion and at each SOA were computed by averaging across subjects. To obtain these CDFs, we rank ordered the RTs in each condition for a given subject (for the SOAs of −100 ms and 100 ms, after adding 100 ms to the RTs in the single-channel condition of the second signal). If any of the 20 RTs were missing, a full distribution was generated by means of a damped cubic SPLINE function. Each of the 20 ordered RTs estimated the RTs at the 2.5th, 7.5th, and 12.5th–97.5th percentiles of the true CDF for a given subject. Composite CDFs were then formed by averaging, across subjects, all the RTs for a given percentile (Miller, 1982).

The redundant signal CDFs were compared with the sum of the single-channel CDFs, represented by Inequalities A3 and A4 in the Appendix. For the SOA of −100 ms, Inequality A4 was violated throughout the range from the 2.5th to the 37.5th percentiles of RT for the Aa condition and in percentiles 27.5 and 32.5 for the Uu condition, as demonstrated by paired t tests between redundant signal and sum-curve distributions across subjects at each of the 20 percentile points in a session. It was never significantly violated under this SOA for the Ua and Au conditions. For the SOA of 0 ms, Inequality A3 was violated from the 2.5th to the 47.5th percentile for Condition Aa, from the 2.5th to the 37.5th percentile of RT for Condition Uu, from the 2.5th to the 17.5th percentile of RT for Condition Au, and from the 2.5th to the 17.5th percentile of RT for Condition Uu.

For the SOA of 100 ms, some puzzling results were obtained. Inequality A4 was never significantly violated for Condition Aa, but it was significantly violated for Condition Au from the 2.5th to the 42.5th percentile, for Condition Ua for the 7.5th and 12.5 percentiles, and for Condition Uu from the 7.5th through 17.5th percentile, staying marginally significant over a longer range (e.g., at the 37.5th percentile, $p = .08$).

After these analyses of the adapted redundant conditions,

an ANOVA was conducted on the unadapted bimodal nonredundant conditions. This analysis showed significant main effects for condition, $F(7, 210) = 26.74$, $p < .001$, and for SOA, $F(2, 60) = 5.32$, $p < .01$, and a significant Condition × SOA interaction, $F(14, 420) = 9.61$, $p < .001$. The mean RTs over SOA for all bimodal nonredundant conditions are presented in Table 7. For each SOA, we tested all bimodal conditions with varying visual or auditory targets against each other and against the single-channel conditions by Newman-Keuls analyses with an alpha of .05. Of these 15 comparisons for each SOA, the comparisons between conditions with the same target are presented in Table 8.

Discussion

When corrected for the effects of statistical facilitation, the results confirmed and extended those of the SOA of −100 ms (visual stimulus led by 100 ms) and the SOA of 0 ms (simultaneous presentation) in Experiment 1. Again, the RTs in congruent redundant conditions (Aa and Uu) showed a much larger facilitation effect than the incongruent ones (Au and Ua) after the RTs were corrected for stimulus differences by subtracting the means predicted on the basis of independent separate activation. Furthermore, the range over which there were violations of the independent activation assumption was much larger for the congruent conditions than for the incongruent conditions. Because the results with mixed stimulus presentation replicate those with blocked presentation, RT differences between the single-channel conditions of Experiment 1 cannot have contributed to the observed differences between congruent and incongruent redundant conditions. Instead, the similarity of results provides strong support for the hypothesis of automatic cross-modal effects at the representation level that occur regardless of changes in the experimental design.

The mixed design of Experiment 2 allowed us to compare not only congruent and incongruent conditions that shared their auditory stimulus (e.g., Aa and Ua, as in Experiment 1), but also those that had their visual stimulus in common (e.g., Aa and Au). For the SOA of −100 ms (in which the visual stimulus started 100 ms before the auditory one), congruent

Table 7
Mean Reaction Times (in Milliseconds) in the Nonredundant and Single-Channel Conditions at Three Stimulus-Onset Asynchronies (SOAs)

| Condition | Single-channel visual | SOA | | | Single-channel auditory |
		−100 ms	0 ms	100 ms	
Ai	406	407	419	417	
Ui	422	428	426	437	
An	406	386	376	380	
Un	422	402	390	395	
Ia		407	404	411	396
Iu		440	448	421	443
*a		386	401	409	396
*u		413	441	432	443

Note. Reaction times were measured from the target stimulus.

Table 8
Newman-Keuls Analyses Testing Bimodal Nonredundant Conditions With Visual and Auditory Targets Against Each Other and Against Single-Channel Conditions at Three Stimulus Onset Asynchronies (SOAs)

| Target | SOA | | |
	−100 ms	0 ms	100 ms
Visual			
Ai vs. An	>	>	>
Ui vs. Un	>	>	>
Ai vs. A−	—	>	>
Ui vs. U−	—	—	>
An vs. A−	<	<	<
Un vs. U−	<	<	<
Auditory			
Ia vs. *a	>	—	—
Iu vs. *u	>	—	—
Ia vs. −a	—	—	—
Iu vs. −u	—	—	<
*a vs. −a	—	—	—
*u vs. −u	<	—	—

Note. Alpha was set at .05. The direction of significant differences is indicated; for example, > indicates that the first condition was slower than the second. Short dashes indicate no signal; long dashes indicate no significant difference.

conditions should be compared with incongruent conditions with the same visual target, because at this SOA most responses will be given to the visual modality. Significant RT differences were indeed obtained, confirming and expanding the results of Experiment 1.

Following the same reasoning, under the SOA of 100 ms (the auditory signal led by 100 ms), the RTs for Condition Aa should be compared with those for Condition Ua and the RTs for Condition Uu with those for Condition Au. These pairs of redundant conditions had their auditory target stimulus in common, which would be expected to elicit the most reactions at this SOA. Let us consider first the obtained pattern of results for the Aa condition, and then that for the Uu condition. After correcting for the characteristics of the contributing single channels, we observed a larger facilitation effect in the congruent Aa' condition than in the incongruent Ua' condition: 15 ms for Aa' versus −23 ms for Ua' (see Table 6). This indicates that for the Aa condition the auditory phoneme /a/ was cross-modally activated by the corresponding grapheme A. The results of the coactivation analysis are not inconsistent with this interpretation of the results. However, they did not allow a rejection of an explanation of the facilitation effect in the Aa condition in terms of statistical facilitation. The pattern of results for the congruent Uu' condition was quite different. This condition showed a slightly larger facilitation effect than the incongruent Au' condition, but this difference did not reach statistical significance (25 ms for Uu' vs. 17 ms for Au'). However, significant coactivation effects were demonstrated in this condition, indicating cross-modal effects at some processing level.

To summarize, the results for the Aa condition showed a significant RT facilitation effect but did not lead to a rejection of an interpretation in terms of only statistical facilitation, whereas the results for the Uu condition did not show sig-

nificant facilitation effects but led to a rejection of a race model without coactivation. As a whole, these results can be interpreted as tentative evidence in favor of grapheme-to-phoneme activation. In combination, the finding of facilitation effects at both the SOA of −100 ms (visual signal led by 100 ms) and the SOA of 100 ms (auditory signal led by 100 ms) supports bidirectional cross-modal activation spreading. Because the experimental results for the 100-ms SOA seem less solid than those for the −100-ms SOA we conducted an additional experiment, intended to replicate Experiment 2, to obtain further support for this conclusion.

Before introducing Experiment 3, we wish to present the results for the nonredundant bimodal conditions. When the RTs in the redundant (Table 5) and nonredundant conditions (Table 7) were compared, it was immediately clear that the nonredundant conditions were much less sensitive to the SOA manipulation than were the redundant conditions. Whereas the redundant conditions consisted of two target stimuli in different modalities, the nonredundant consisted of a target stimulus in one modality and a neutral stimulus in the other. The insensitivity of the nonredundant conditions to SOA was therefore to be expected, because the neutral stimulus should have only indirect effects on the responding (e.g., via arousal): No effects of statistical facilitation can occur when only one target is involved.

By incorporating two different types of nonredundant trials, we tried to bridge the gap between studies on the bimodal processing of nonlinguistic signals and those involving linguistic material. As can be seen in Table 7, the RT differences between nonredundant conditions with auditory signal /i/ and those with a white-noise signal were larger when these neutral stimuli arrived earlier with respect to the visual target. Whereas the white-noise signal had general facilitatory effects with respect to the single-channel conditions, the effect of /i/ was slightly inhibitory. As Table 7 indicates, there were RT differences between nonredundant conditions with visual nontargets I and * as well, but only when these neutral stimuli preceded the auditory targets.

Overall, the effects of a neutral stimulus on the target seemed to be stronger and more stimulus dependent when the neutral stimulus was auditory than when it was visual. The following account is in line with the general pattern of results. Auditory stimuli characteristically cause relatively larger arousal effects than do visual stimuli (Keuss, 1987). Therefore, larger facilitation effects with respect to the single channel may be expected for the white-noise conditions than for the * conditions. However, to reject a stimulus as a possible target, a subject must at least perform a partial analysis of it. It seems plausible that the amount of analysis required depends on the stimulus's similarity to a target. Because white noise is very different from speech, the white-noise signal can quickly be rejected as a possible target. Relative to white noise, a star stimulus will take longer to reject, because there is less that distinguishes it from a target letter. If the stimulus is the letter I or speech sound /i/, it may be still harder to reject as a target. According to this view, interference occurs in the "I"-conditions because the neutral stimulus must be processed to a certain extent before the attention can be shifted to the other channel. This interference effect

hides the general facilitation caused by an auditory stimulus in case of the neutral /i/.

These results suggest that two components should be distinguished in the effect of linguistic stimuli: both a general nonlinguistic component and a component that depends on the linguistic identity of the stimulus. In processing models, the relative importance of both effects should be considered with respect to the visual and auditory stimuli that are involved.

We now return to the important issue of grapheme-to-phoneme activation at an SOA of 100 ms. Although most results were in accordance with the hypothesis that the auditory target stimulus is influenced by the visual stimulus following at a delay of 100 ms, one would expect stronger cross-modal effects on the basis of the literature. For example, the existence of sublexical grapheme-to-phoneme activation has been discussed and supported in a review by Van Orden et al. (1990) and in the studies by Dijkstra et al. (1989) and Perfetti and his colleagues (Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988). The hypothesis that grapheme–phoneme interaction is bidirectional would be strengthened considerably if the partially positive results of Experiment 2 could be replicated and confirmed. We therefore decided to run another experiment, in which only the most essential conditions and SOAs were incorporated. To test the generality of the cross-modal representation effects obtained previously, we changed the target set to a different stimulus combination: from A and U to A and E.

Experiment 3

Method

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

Design. In this experiment, only four bimodal redundant conditions (Aa, Ae, Ea, and Ee) and four single-channel conditions (A–, E–, –a, –e) were included in the go stimulus set. The no-go set was analogously organized with stimuli O and U. In the bimodal trials two SOAs were used, presenting the visual stimulus either 100 ms before or 100 ms after the onset of the auditory stimulus. The first of the two signals was presented 500 ms after the warning signal. In the single-channel conditions, the visual or auditory signal was presented 500 or 600 ms after the warning signal, in correspondence with the two SOAs of the redundant conditions. All visual stimuli were displayed for 320 ms, the average duration of the auditory target stimuli.

The number of repetitions for each SOA was 20 in the redundant conditions and 12 in the single-channel conditions. Two sessions were conducted, each consisting of a total of 512 test stimuli subdivided into four separately randomized presentation blocks: 2 (go/no go) × 4 (blocks) × 4 (conditions) × 2 (SOA) × 5 (repetitions per block) + 2 (go/no go) × 4 (blocks) × 2 (conditions) × 2 (modality) × 2 (SOA) × 3 (repetitions) = 320 + 192 = 512.

Each presentation block was thus constructed in such a way that it could be analyzed separately, in order to examine the development of order effects over the experiment. An additional 40 practice trials were constructed, leading to a total number of 552 trials in one session.

Stimuli. The auditory stimuli were identical to those used in Experiments 1 and 2. However, in this experiment the speech stimuli were digitally conserved and retrieved from hard disk at the moment of presentation. Presentation was binaural, as before. The visual stimuli consisted of white Times Roman capital letters, 12 mm in height, presented on a NEC-Multisync color monitor using a VGA 1024 graphic display adapter card. Presentation of the visual and auditory stimuli, as well as the registration of the reaction times, was controlled by the experimental software of the Max-Planck NESU-System, consisting of a host personal computer (IBM-386), connected to and synchronized with a timer server (IBM-XT with 1-ms reliability) and a speech server (IBM-XT).

Procedure. The procedure was analogous to that in earlier experiments, except that the subjects were now instructed to react as soon as possible if they detected the letter and/or speech sound "A" or "E" (Dutch).

Each trial started with a 1000-Hz warning signal with a duration of 200 ms. In the single-channel conditions, this signal was followed after 500 or 600 ms by either a visual or an auditory stimulus. In the redundant conditions, a visual or an auditory stimulus appeared after 500 ms, followed after a delay of 100 ms by an auditory or visual stimulus, respectively. Two seconds after presentation of the last signal, a new trial started.

After the practice set of 40 trials there was a short break during which questions could be posed. A session of 552 trials took about 40 min, including three short breaks of 1 min, after each block of 128 trials.

Results

The data from each of the two sessions were analyzed separately. Because floor effects clearly affected the results for Session 2, only the results of Session 1 are reported.[6] Mean RTs (measured from the onset of the first presented target stimulus) were computed for each subject and each experimental condition. Latencies longer than 750 ms or shorter than 150 ms were treated as errors. The total percentage of missing and extreme values was 1.6%. Errors were substituted by mean RTs in each subcondition for each subject. The percentage of false alarms (i.e., reactions to the no-go trials) was 2.0%. Table 9 shows the raw mean results in the single-channel conditions of Session 1 with stimulus signals presented 500 or 600 ms after the warning signal, as

Table 9
Mean Reaction Times (in Milliseconds) for the Single-Channel Conditions When the Stimulus Signal Was Presented 500 ms or 600 ms After the Warning Signal Ended

	Signal		
Condition	500 ms after warning	600 ms after warning	All[a]
A–	389	387	388
E–	390	391	391
–a	421	421	421
–e	438	446	443

Note. Dashes indicate no signal in either the auditory (e.g., A–) or visual (e.g., –a) modality.
[a] Average of reaction times to 500-ms and 600-ms stimulus signals combined.

Table 10
Mean Reaction Times (in Milliseconds) Predicted by an Independent Separate Activation Model and Obtained for the Redundant Conditions Aa, Ea, Ae, and Ee at Two Stimulus Onset Asynchronies (SOAs) and t Tests and Correlations for the Predicted and Obtained Means

Condition	Predicted	Obtained	Difference	t(32)	p	r
	SOA = –100 ms (letter led by 100 ms)					
Aa	381	379	2	–.55	ns	.93
Ae	383	393	–10	3.18	<.01	.95
Ea	384	390	–6	1.35	ns	.92
Ee	387	384	3	–.79	ns	.93
	SOA = 100 ms (speech sound led by 100 ms)					
Aa	399	385	14	–2.78	<.01	.91
Ea	399	408	–9	2.18	<.05	.92
Ae	415	424	–9	2.40	<.05	.95
Ee	417	395	22	–5.21	<.001	.91

Note. All correlations significant at $p < .001$.

well as the overall mean RTs for those conditions.[7] The obtained means of the bimodal conditions are shown in Table 10 for each of the two SOAs separately. The block factor is ignored in Table 10.

Independent race predictions. Table 10 also contains the means predicted on the basis of an independent race model. We computed the predicted minimum distribution for each subject and each condition by using the single-channel conditions according to Equation A1 in the Appendix. After adding 100 ms to the single-channel RTs of the later signal, we ordered the correct RTs in millisecond steps for the analysis. Only RTs between 150 and 750 ms were included. For each SOA, the mean predicted bimodal RTs thus computed were subtracted from the corresponding empirically obtained RTs. Table 10 shows the size and significance of these differences between the predicted and obtained RTs for the four redundant conditions at the two SOAs.

---

[6] Analysis of Session 2 showed that subjects were so well practiced that the average RTs in most conditions fell below 375 ms. For those conditions, the differences between adapted congruent and incongruent conditions did not reach statistical significance, although all trends were in the expected direction. For conditions in which RTs were slower than 375 ms, significant results similar to those found in Session 1 were obtained.

[7] As can be seen in Table 9, the mean single-channel RTs for the visual stimuli A and E were practically identical. Assuming that the processing of these stimuli during bimodal trials of which they are a component is similar as well, one can compare congruent and incongruent conditions for an SOA of 100 ms without applying the independent race correction. For this SOA, the statistical facilitations caused by the visual stimuli A and E may be assumed to be approximately equal. Planned comparisons show that the difference between the raw means in Aa and Ea conditions was highly significant (385 vs. 408 = 23 ms), $t(32) = -4.16, p < .001$, as was the difference between Ee and Ae conditions (395 vs. 424 = 29 ms), $t(32) = -8.35, p < .001$.

Table 11

Size and Significance of Reaction Time Differences (in Milliseconds) Between Adapted Congruent and Corresponding Incongruent Redundant Conditions at Two Stimulus Onset Asynchronies (SOAs)

Congruent	Incongruent	Difference	$t(32)$	p
SOA = −100 ms (letter led by 100 ms)				
Aa′	Ae′	12	−3.99	<.001
Ee′	Ea′	9	−1.96	.06
SOA = 100 ms (speech sound led by 100 ms)				
Aa′	Ea′	23	−4.11	<.001
Ee′	Ae′	31	−8.48	<.001

An ANOVA was conducted using the adapted redundant conditions and showed significant main effects of condition, $F(3, 96) = 19.10$, $p < .0001$, and SOA, $F(1, 32) = 4.35$, $p < .05$, as well as a significant Condition × SOA interaction, $F(3, 96) = 4.70$, $p < .01$. For the adapted data, planned paired comparisons were conducted (alpha = .05), showing significant differences between congruent and corresponding incongruent conditions, as indicated in Table 11. All comparisons were significant, except for the comparison of the Ea′ and Ee′ conditions for the SOA of −100 ms, which was only marginally significant ($p = .06$).

Coactivation test. Subsequently, Miller's (1982, 1986) test for the presence of coactivation was conducted. As before, the average CDFs for the single-channel visual and auditory conditions (disregarding SOA) and the redundant conditions for each SOA were computed by averaging across subjects.

To obtain these CDFs, we rank ordered the RTs in each condition for a given subject (after adding 100 ms to the RTs in the single-channel condition of the second signal). The maximum number of RTs was thus 20 in the redundant conditions and 24 in the single-channel conditions (the number of replications disregarding SOA allocation). A full distribution, based on 24 replications, was generated by means of a damped cubic SPLINE function. Each of the 24 ordered RTs estimated the RTs at the 2.08th, 4.17th, and so forth percentiles of the true CDFs for a given subject. Composite CDFs were then formed by averaging, across subjects, all of the RTs for a given percentile.

The redundant-signal CDFs were compared with the sum of the single-channel CDFs, represented by Inequalities A3 and A4 in the Appendix. For the SOA of −100 ms, Inequality A4 was violated throughout the range from the 2.08th to the 18.76th percentiles of RT for the Aa condition and from the 6.25th to the 22.93th percentiles for the Ee condition. For the SOA of 100 ms, Inequality A4 was violated throughout the range from the 10.42th to the 64.63th percentile for the Aa and from the 2.08th to the 64.63th percentile for the Ee, as demonstrated by paired t tests between redundant-signal and sum-curve distributions across subjects at each of the 24 percentile points. It was never significantly violated at either SOA for the incongruent Ae and Ea conditions. The results thus clearly indicate the presence of coactivation in the congruent redundant

conditions, while a separate activation model cannot be rejected in the case of the incongruent conditions.

To determine whether the effects of congruence remained constant over the experimental session, for each subject and block we computed the RT differences between each redundant condition and the corresponding single-channel condition of the first presented stimulus (e.g., between Aa and A−). Table 12 shows these RT differences for all four redundant conditions and both SOAs. We then performed an ANOVA on these data with the factors congruence, block and SOA. Significant main effects were found for congruence, $F(1, 32) = 18.53$, $p < .001$, and SOA, $F(1, 32) = 53.99$, $p < .001$. The only significant interaction was between congruence and SOA, $F(1, 32) = 20.78$, $p < .001$. A further analysis of the data for the SOA of −100 ms separately resulted only in a significant main effect of congruence, $F(1, 32) = 15.33$, $p < .001$. Thus, the congruent conditions showed consistent larger RT differences with respect to the single-channel conditions over blocks than did the incongruent conditions, for both SOAs.

Discussion

Experiment 3 replicates and strengthens the results of Experiments 1 and 2 for both the −100-ms SOA (visual target stimulus preceded auditory) and the 100-ms SOA (auditory target preceded visual). For both SOAs and practically all conditions, significant RT advantages of the adapted congruent over the incongruent conditions were obtained. Moreover, these differences did not change significantly in size over successive presentation blocks. This finding suggests that the congruency effect is not attributable to strategy effects, which are generally assumed to evolve over the course of the experiment. Finally, the results of Miller's (1982, 1986) tests are completely consistent with these conclusions, because by falsifying separate activation models they indicate the presence of coactivation in the congruent conditions,

Table 12

Reaction Time Differences (in Milliseconds) Between the Redundant and Single-Channel Conditions With the Same First Signal for Each of Four Blocks and Two Stimulus Onset Asynchronies (SOAs)

Condition	Block			
	1	2	3	4
SOA = −100 ms				
Aa/A−	2	25	9	4
Ae/A−	−11	2	−7	2
Ea/E−	−5	11	3	−9
Ee/E−	5	5	6	6
SOA = 100 ms				
Aa/−a	35	39	38	32
Ea/−a	5	11	23	14
Ae/−e	12	21	5	19
Ee/−e	46	42	42	44

Note. The first letter of a condition refers to the visual stimulus, the second to the auditory; a dash indicates no signal.

but do not reject separate activation models in the incongruent conditions.

General Discussion

In the present experiments, we examined the relationship between grapheme and phoneme representations by means of a bimodal detection task, in which subjects reacted to specific target letters (or symbols) and/or speech sounds. This type of task is well-suited for investigation of automatic cross-modal activation effects because it yields fast RTs and allows for, but does not require, on-line cross-modal effects. The task was used to examine three aspects of the grapheme–phoneme relationship. We considered first the existence of cross-modal facilitation effects and, second, the size and directionality of such effects. Finally, we tested for cross-modal inhibition effects.

The activation of graphemes by phonemes and vice versa was demonstrated in all three experiments. After being corrected for the occurrence of stimulus-dependent statistical facilitation effects, the RTs in congruent redundant conditions (e.g., Aa) were facilitated compared with those in incongruent conditions (e.g., Au). This facilitation was obtained when the visual stimulus preceded the auditory one by 100 ms, when it followed the auditory signal by 100 ms, and when both stimuli were presented simultaneously. Because the congruent and incongruent conditions were comparable in all respects except the nominal identity of their component stimuli, the relative facilitation in the congruent conditions must be ascribed to cross-modal activation at a representation level. This conclusion was further supported by the presence of strong coactivation effects in the congruent conditions.

Thus, the observed facilitation effects can be localized at the representation level by the combination of two methodological innovations derived from research on nonlinguistic bimodal processing: the prediction of RTs on the basis of a race model and the application of Miller's (1982, 1986) test for the presence of coactivation.

The present experiments not only provide evidence that the facilitation effects were representational in nature, but also indicate that these effects were automatic (Posner & Snyder, 1975). First, the facilitation effects were unlikely to have been under subjects' strategic control, because consistent effects were found in experiments with mixed and blocked designs and with different instructions and stimulus conditions. Second, the facilitation effects were fast, because they were even obtained in RTs of about 330 ms under an SOA of 100 ms (Experiment 1).

The second issue that was investigated was whether cross-modal influences from the visual to the auditory modality are similar in size and time course to those in the other direction. In the bimodal redundant conditions, fast cross-modal activation effects were found to occur both from the auditory to the visual modality (letter led by 100 ms) and from the visual to the auditory domain (letter followed by 100 ms). Given that the visual and auditory single-channel RTs did not differ much, the results for the first SOA can be considered evidence in favor of the existence of phoneme-to-grapheme ac-

tivation, and those for the latter SOA support grapheme-to-phoneme activation effects.

The existence of grapheme-to-phoneme activation effects was evident from the significant facilitation effects obtained in Experiments 2 and 3 when the auditory stimulus preceded the visual one. The conclusion of automatic grapheme-to-phoneme activation fits well with recent results obtained with the backward masking and masked priming techniques in the area of visual word recognition (Perfetti & Bell, 1991; Perfetti, Bell, & Delaney, 1988). Perfetti et al. (1988) asked subjects to identify briefly presented lowercase target words that were followed first by an uppercase pseudoword mask and subsequently by a pattern mask (a row of Xs). The orthographic and phonological properties shared by the target word and pseudoword mask were varied. When homophonic (MAYD) and orthographically similar (MARD) masks were equated for number of letters shared with the target word (made), both conditions led to a higher percentage of correct identifications of the target than a control mask, but an additional improvement of performance was found for the homophonic mask over the orthographically similar mask. The authors ascribed this effect to automatic "phonetic activation" and concluded 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.

Perfetti and Bell (1991) turned the backward masking task into a masked priming situation by presenting the pseudoword before the target and varied the exposure duration of the masked pseudoword prime. When the prime was shown for 45 ms or longer, a phonemic effect began to emerge in target word identification. This again supported the conclusion that phonemic activation can occur prior to the identification of a printed word.

Phoneme-to-grapheme activation effects were found in all experiments for an SOA of −100 ms, at which the visual stimulus preceded the auditory one by 100 ms. Phoneme-to-grapheme activation effects have not been investigated in word recognition as much as grapheme-to-phoneme activation effects have been, but the available studies are again consistent with our findings (Donnenwerth-Nolan et al., 1981; Seidenberg & Tanenhaus, 1979; Tanenhaus et al., 1980). For example, in cross-modal research Tanenhaus et al. (1980) observed that color-naming latencies were longer (compared with control conditions) when the target words were preceded by auditory prime words that were phonologically or orthographically similar. Our results add substantially to these earlier studies in that they indicate, first, that phoneme-to-grapheme effects do not require mediation by the lexical (word) level and, second, that such effects can arise automatically in a detection task in which (in contrast to a naming task) an orthographic representation would be sufficient to react to the visual target stimulus.

The third issue we investigated concerned cross-modal inhibition effects. Whereas cross-modal facilitation effects between associated sublexical units were clearly present in the redundant conditions, no evidence was collected in favor of

cross-modal inhibition or of mediated inhibition at a representation level (e.g., grapheme A inhibiting phoneme /u/ via activation of phoneme /a/). Incongruent redundant conditions involving two-letter targets (e.g., Ua) showed RT patterns that were very similar to those found for conditions combining an * with a letter. It could be argued that this symbol was in some respects similar to a letter and thus is itself capable of cross-modal inhibitory influences. If this were so, however, one would still expect a dependence of the size of the inhibitory effect on SOA (i.e., more inhibition for the SOA of 0 ms, where distributions overlapped more), which was not observed.

Because there are few bimodal studies available that address the issue of inhibition across modalities (but see Dijkstra et al., 1989), further evidence is needed to corroborate this finding. However, the present data already impose strong constraints on the possible architecture of the human language processing system. They indicate that the connectionist models were right not to introduce inhibition effects between visual and auditory processing modules. Future research should also address the issue of inhibition between representations within one modality, because many connectionist models do assume such inhibition (e.g., the TRACE model by McClelland & Elman, 1986).

Taken together, the results of our experiments indicate the existence of automatic bidirectional activation and the absence of cross-modal inhibition between graphemes and phonemes. This conclusion should be of interest to all researchers whose models address these kinds of representations. Because the cross-modal effects were shown to be fast and presumably automatic, the results provide some support for the hypothesis that grapheme representations play a role in auditory word recognition, and phoneme representations in visual word recognition. This suggestion would be in line with a type of dual route model (Coltheart, Davelaar, Jonasson, & Besner, 1977) or time course model (Seidenberg, 1985a, 1985b) for the auditory modality. An interesting consequence of this viewpoint would be that the recognition of an auditory target word would be influenced not only by the set of similar words within the same modality (its auditory cohort), but also by the visual counterpart of that set of word candidates.

It is interesting to show how our results fit in with a model of visual word recognition like that of Seidenberg and McClelland (1989). This model makes several predictions that agree with our data. For example, it predicts cross-modal grapheme-to-phoneme facilitation effects, but no inhibition effects. When a provision is made to allow auditory input to the model and when phoneme-to-grapheme connections are realized, it would also be in line with our finding of bidirectional cross-modal activation. Indeed, Seidenberg and McClelland themselves have suggested that incorporating phoneme-to-grapheme effects would solve some problems of the model with respect to pseudohomophone effects (Seidenberg & McClelland, 1990). Our results could be interpreted likewise within the word recognition accounts developed by Lukatela et al. (1989; Lukatela et al., 1990) and by Van Orden et al. (1990).

All these models mainly address lexical aspects of language processing. As described in the Introduction, diverse models on nonlinguistic bimodal processing (e.g., several types of coactivation models) are available as well. One major challenge is to develop new models for combined visual and auditory processing that can handle both linguistic and nonlinguistic stimuli. Some recent models already seem to hold a promise for such future developments (Dijkstra, 1990; Mordkoff & Yantis, 1991; Van Orden et al., 1990).

References

Appelman, I. B., & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics, 30,* 436–446.

Colonius, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology, 34,* 253–275.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Craig, A., Colquhoun, W. P., & Corcoran, D. W. J. (1976). Combining evidence presented simultaneously to the eye and the ear: A comparison of some predictive models. *Perception & Psychophysics, 19,* 473–484.

De Boor, C. (1978). *A practical guide to splines.* New York: Springer-Verlag.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research, 49,* 23–29.

Dijkstra, A. (1990). *Cross-modal contacts between graphemes and phonemes. Explorations in bimodal processing* (NICI Technical Report 90-02). Katholieke Universiteit Nijmegen, Nijmegen, The Netherlands.

Dijkstra, A., Schreuder, R., & Frauenfelder, U. H. (1989). Grapheme context effects on phonemic processing. *Language and Speech, 32,* 89–108.

Donnenwerth-Nolan, S., Tanenhaus, M. K., & Seidenberg, M. S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 170–180.

Ehri, L. C. (1985). Effects of printed language acquisition on speech. In D. R. Olson, N. Torrance, & A. Hildegard (Eds.), *Literacy, language and learning: The nature and consequences of reading and writing* (pp. 333–367). Cambridge, England: Cambridge University Press.

Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics, 44,* 191–193.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory & Cognition, 17,* 302–310.

Frost, R., Repp, B. H., & Katz, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory and Language, 27,* 741–755.

Gielen, S. C. A. M., Schmidt, R. A., & Van den Heuvel, P. J. M. (1983). On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics, 34,* 161–168.

James, W. (1950). *The principles of psychology* (Vol. 1). New York: Dover. (Original work published 1890).

Keuss, P. J. G. (1987). Intersensorische facilitatie [Intersensory facilitation]. In J. J. Beishuizen, L. de Leeuw, & J. H. de Swart (Eds.), *Cognitieve psychologie in theorie en praktijk* (pp. 39–59). Lisse, The Netherlands: Swets & Zeitlinger.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* London: Oxford University Press.

Lukatela, G., Carello, C., & Turvey, M. T. (1990). Phonemic priming with words and pseudowords. *European Journal of Cognitive Psychology, 2,* 375–394.

Lukatela, G., Turvey, M. T., Feldman, L. B., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetical word perception. *Journal of Memory and Language, 28,* 214–236.

MacKay, W. A., & Bonnet, M. (1990). CNV, stretch reflex and reaction time correlates of preparation for movement direction and force. *Electroencephalography and Clinical Neurophysiology, 76,* 47–62.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279.

Miller, J. (1985). Discrete and continuous models of divided attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 513–528). Hillsdale, NJ: Erlbaum.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40,* 331–343.

Miller, J. (1991). Channel interaction and the redundant target effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 160–169.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 520–538.

Mulligan, R. M., & Shaw, M. L. (1980). Multimodal signal detection: Independent decisions vs. integration. *Perception & Psychophysics, 28,* 471–478.

Mullin, P. A., Egeth, H. E., & Mordkoff, J. T. (1988). Redundant-target detection and processing capacity: The problem of positional preferences. *Perception & Psychophysics, 43,* 607–610.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review, 80,* 489–509.

Perfetti, C. A., & Bell, L. C. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30,* 473–485.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59–70.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Raab, D. H. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York Academy of Sciences, 24,* 574–590.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446–461.

Scinto, L. F. M. (1986). *Written language and psychological development.* San Diego, CA: Academic Press.

Seidenberg, M. S. (1985a). The time course of information activation and utilisation in visual word recognition. In D. Besner, T. G. Waller, & G. E. Mackinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 199–249). San Diego, CA: Academic Press.

Seidenberg, M. S. (1985b). The time course of phonological code activation in two writing systems. *Cognition, 19,* 1–30.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., & McClelland, J. L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review, 97,* 447–452.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546–554.

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological priming in auditory and visual word recognition. *Memory & Cognition, 8,* 513–520.

Ulrich, R., & Giray, M. (1986). Separate-activation models with variable base times: Testability and checking of cross-channel dependency. *Perception & Psychophysics, 39,* 248–254.

van der Heijden, A. H. C., Schreuder, R., Maris, L., & Neerincx, M. (1984). Some evidence for positively correlated separate activation in a simple letter-detection task. *Perception & Psychophysics, 36,* 577–585.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

Wickelgren, W. A. (1969). Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76,* 1–15.

Wood, L. E. (1974). Visual and auditory coding in a memory matching task. *Journal of Experimental Psychology, 102,* 106–113.

Wood, L. E. (1977). Modality-specific coding in matching letters. *Perception & Psychophysics, 21,* 329–335.

Appendix

Method of Analysis

To obtain an indication of the size of the statistical facilitation effect in a particular redundant condition, and to correct for this effect, one needs to predict the reaction time (RT) for the redundant condition on the basis of the single-channel RTs. For this prediction, techniques developed by Miller (1982, 1986) and others are indispensable, and their rationale is explained as follows.

The most simple type of separate activation model assumes independent channels. For such a model, processing in a redundant trial may be likened to a horse race. If two horses, V and A, race, the chance that the race is won by one of the two horses at time t is equal to the chance that horse V has finished at that time plus the chance that horse A has finished at that time, minus the chance that both have finished; in mathematical terms,

$$P(RT_{va} \leq t) = P(RT_v \leq t) + P(RT_a \leq t)$$
$$- P(RT_v \leq t \wedge RT_a \leq t). \qquad (A1)$$

Here, with independent channels, the last term is equal to the product $P(RT_v \leq t)P(RT_a \leq t)$. If reactions to redundant trials are considered the result of such a race between the visual and auditory channels, Equation A1 can be used to derive the predicted minimum distribution resulting from independent auditory and visual single-channel distributions.

To compare the information about the distributional characteristics of the RTs in the obtained and predicted conditions and to test separate activation models in general against coactivation models, Miller (1982) used the fact that the last term in Equation A1 is always equal to or greater than 0, that is,

$$P(RT_v \leq t \wedge RT_a \leq t) \geq 0. \qquad (A2)$$

It follows that with separate activation, for all values of t,

$$P(RT_{va} \leq t) \leq P(RT_v \leq t) + P(RT_a \leq t). \qquad (A3)$$

If this inequality is violated, all separate activation models (whether dependent or independent) have to be rejected; that is, when the (estimated) probability of occurrence of latencies smaller than some value t in the redundant condition exceeds the sum of the (estimated) probabilities in the two single-channel conditions, separate activation models are rejected. Coactivation models are consistent with violation of Inequality A3, because, with pooling of activation, the fastest responses to redundant signals can be faster than the fastest response to each channel alone.

To evaluate separate activation models under signal conditions in which one signal precedes the other by a certain time lag, Miller (1986) extended Inequality 3. According to such models, a response to redundant signals is caused by the first of the two separate processes responding to each signal to finish. If signal presentation is asynchronous, the two processes do not start at the same time and the finishing times must be adjusted to take that fact into account. If RT is measured from the onset of the first signal, the stimulus onset asynchrony (SOA) between the two signals must be added to the latency of the responses to the sec-

ond signal. Under such circumstances the following inequality should hold:

$$P(RT_{va} \leq t) \leq P(RT_v \leq (t - SOA_v)) + P(RT_a \leq (t - SOA_a)) \qquad (A4)$$
$$\text{for all } t.$$

In Inequality A4, SOA_v and SOA_a denote the SOAs from the onset of the first signal to the onset of the visual and auditory signals, respectively. Either SOA_v or SOA_a will be 0 in a particular redundant trial. When the visual signal precedes the auditory by an SOA_a, for example, SOA_v will be 0.

The equations just described were used in the analysis of the present experimental data in the following way. Assuming that the contributions of decision and motor level coactivation are comparable in the congruent (e.g., Aa) and incongruent (e.g., Ua) conditions, we expected the extra contribution of coactivation at a representation level in the congruent condition to lead to more coactivation overall. Whether more coactivation was indeed present in the congruent than in the incongruent conditions was determined by comparing the obtained RTs in those conditions after they were corrected by subtracting the RTs predicted by a simple race model, which assumes independent and uncorrelated channels.[A1]

During data analysis, we used Equation 1 to compute for a specific bimodal condition and subject the expected mean if the race model with independent channels would hold (using the obtained visual and auditory single-channel data as estimators for $P(RT_v \leq t)$ and $P(RT_a \leq t)$). This mean was subsequently subtracted from every RT for that condition and subject, and finally the corrected distributions were tested against each other.

To test whether coactivation was in fact present in the experimental condition and SOA at hand, we applied Inequality A4 to determine whether the obtained data distribution led to a rejection of separate activation models in general. A violation of Inequality A4 provided additional support for the presence of coactivation in that condition.

[A1] According to Mulligan and Shaw (1980), "there is agreement that auditory and visual processing are not correlated in a bimodal detection task" (p. 472). However, with respect to letter detection tasks, a negative dependence between processing channels is often reported in the experimental literature (e.g., van der Heijden, Schreuder, Maris, & Neerincx, 1984). It is therefore not completely clear whether independence can be assumed for our experiments. However, as long as any dependence does not differ systematically between conditions (and there does not seem any reason for that), our approach is still valid for comparing differences between conditions. Moreover, we also tested differences in the amount of coactivation between congruent and incongruent conditions by means of an estimation of the surface violation present in both cases (following Miller, 1986). The results of these tests always led to the same conclusions as we present in this article, which were based on the assumption of independent channels.