Applied Psycholinguistics **23** (2002), 509–533
Printed in the United States of America
DOI: 10.1017.S0142716402004022

The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling

STEPHEN A. BIRD and JOHN N. WILLIAMS
University of Cambridge

ADDRESS FOR CORRESPONDENCE
John N. Williams, Research Centre for English and Applied Linguistics, University of Cambridge, Keynes House, Trumpington Street, Cambridge CB2 1QA, United Kingdom. E-mail: jnw12@cam.ac.uk

ABSTRACT
Two experiments examined the effect of single-modality (sound or text) and bimodal (sound and text) presentation on word learning, as measured by both improvements in spoken word recognition efficiency (long lag repetition priming) and recognition memory. Native and advanced nonnative speakers of English were tested. In Experiment 1 auditory lexical decisions on familiar words were equally primed by prior bimodal and sound-only presentation, whereas there were no priming effects for nonwords. Experiment 2 employed a rhyme judgment task using nonwords. Repetition priming of auditory rhyme judgment decisions was now obtained, and this was greater in the bimodal than the sound-only condition. In both experiments prior bimodal presentation improved recognition memory for spoken words and nonwords compared to single modality presentation. We conclude that simultaneous text presentation can aid novel word learning under certain conditions, as assessed by both explicit and implicit memory tests.

It has been proposed that one way of helping learners of English to comprehend authentic video programs while maintaining a target language learning environment is adding English text subtitles to English videos (Vanderplank, 1988). Identical in format to standard translation subtitles found in many foreign films (e.g., French subtitles for an English film), same-language subtitles (e.g., English subtitles for an English soundtrack), also known as *unilingual* or *intralingual* subtitles (Jung, 1990), can be presented at the bottom of a video screen synchronously with a video soundtrack. Current technology, such as subtitling for those with hearing impairments (see Vanderplank, 1988, for a discussion) and recent language-learning multimedia packages (e.g., “Italia 2000”, Burnage, 1997) now present learners and teachers with the option of using videos that

© 2002 Cambridge University Press 0142-7164/02 $9.50

provide simultaneous transcriptions of the soundtracks (see Davis, 1998, for a summary of available media). Because the subtitles neatly divide the boundaries between words in the speech stream and are unaffected by accent variations and sound degradation that can adversely affect auditory comprehension (Vanderplank, 1993), they may provide easier access to the target language and lead to greater comprehension and learning. But what might students be learning in such a context?

A number of studies have demonstrated some beneficial effects of same-language subtitles (Chung, 1996; Danan, 1992; Garza, 1991; Holobow, Lambert, & Sayegh, 1984; Lambert, Boehler, & Sidoti, 1981; Price, 1983; Vanderplank, 1988, 1990, 1993; and see Borras & Lafayette, 1994, and Jung, 1990, for reviews). In these studies the effects of subtitling were primarily measured by first showing students video segments with or without subtitles and then testing "global" comprehension (i.e., written or oral questions about the video's plot), the meanings of individual words, and sometimes memory for exact phrasing. However, the results of these studies may not be particularly surprising because text may present the easiest path to comprehension, and the auditory input might be ignored without loss of the information required for successfully completing a written test. Subjects could certainly perform both written and oral tests of story line content even if they only read the subtitles. Thus, improved comprehension of a video plot and better retention of phrasing and vocabulary could be due only to good reading comprehension, not improved listening comprehension. This issue is important for teachers and learners because it remains unclear whether subtitles lead to better or worse listening comprehension.

Several researchers have suggested that subtitles may aid spoken word recognition development. For example, Garza (1991, p. 246) argues that subtitles may enhance the learning of foreign languages by helping students to "build their aural comprehension in relation to their reading comprehension." Specifically, Garza argues that a student working with subtitles "will not likely miss the aural cue of a captioned expression the next time he/she encounters it." In a similar study, Borras and Lafayette (1994, p. 70) conclude that same-language subtitling "may help the foreign/second language learner associate the aural and written forms of words more easily and quickly than video without subtitles."

These claims go somewhat beyond the demonstration that same-language subtitles improve the comprehension of a video. They suggest that, having seen the written version of a spoken word, the student will be better able to hear that word when it is presented again. If true, this would suggest that presenting text and sound versions of a word can qualitatively change the phonological representation of the word in the student's mind: that is, the text serves to improve the recognition of that auditorily presented word, even when text is not present in later presentations.

In contrast to previous research on subtitling, we chose to focus primarily on the issue of whether bimodal presentation leads to enhanced learning, looking both at familiar and unfamiliar words. We distinguish two types of learning that may be affected by presentation condition. The first is an implicit form of learning relating to auditory word recognition, and the second is an explicit form of learning relating to word retention.

Most of the subtitling studies described above measured explicit memory. Explicit memory is associated with "intentional or conscious recollection of prior experiences, as assessed by traditional tests of recall and recognition" (Schacter, 1992, p. 244). Explicit memory presumably plays an important role in comprehension in terms of allowing the reader or listener to keep track of the developing discourse representation, characters, and plot. It has also been implicated in vocabulary learning (Ellis, 1994) and rule learning (Robinson, 1995). Learners may work out the meanings of words or the structure of rules by consciously recalling and comparing words or sentences that they have encountered. By contrast, implicit memory is associated with "changes in performance or behaviour that are produced by prior experiences on tests that do not require any intentional or conscious recollection of those experiences" (Schacter, 1992, p. 244). This kind of memory is usually associated with implicit learning and skill development. Implicit learning processes are not just confined to the acquisition of new skills, but can also lead to improvements in the fluency of existing skills, such as the recognition of known words.

In order to examine the effects of subtitling on the implicit and explicit learning of words we employed an adaptation of the repetition priming paradigm in a three-phase design. In the first phase subjects performed a lexical decision task on familiar and unfamiliar words in the auditory modality, the visual modality, or both simultaneously. The latter corresponds to a subtitling condition in which text and sound are presented simultaneously. In the second phase subjects again performed a lexical decision on familiar and unfamiliar words, but this time all stimuli were presented in the auditory modality only. Half of the items were "old," in that they had occurred in Phase 1, and half were "new," in that they had not. By comparing response times to old and new words, it was possible to assess whether bimodal presentation in Phase 1 produced any greater improvement in auditory word recognition performance than auditory presentation alone. Such performance improvements are typically regarded as a form of implicit learning (Roediger & McDermott, 1993) and can be obtained even in the absence of conscious recollection that the item had been presented earlier (e.g., Haist, Musen, & Squire, 1991). For words, this improved performance can be seen as facilitation in processing known word forms; for nonwords, faster reaction times can be regarded as the learning of new word forms. In the third phase of the experiments subjects were again presented with the old and new words from Phase 2, but this time they were required to make an explicit recognition memory decision. This is a test of whether bimodal presentation enhances the explicit memory for words.

Repetition priming for words and nonwords both within and across auditory and visual modalities has been studied extensively (for reviews, see Kirsner & Dunn, 1985; Kirsner, Milech, & Standen, 1983; Monsell, 1985; Roediger & McDermott, 1993; Schacter, 1992). Although no studies have compared the effects of adding a modality to the first presentation in the way we propose here, it is relevant to consider those studies that have examined visual–auditory priming because, if this occurs, it would be suggestive of the kind of interaction between orthographic and phonological codes that is necessary to support an advantage for bimodal presentation. A number of studies have found no priming effect for words and nonwords in visual study/auditory test cross-modal conditions (e.g., Ellis, 1982; Gipson, 1984). Allport and Funnell (1981) argued from such data that repetition priming is a modality-specific phenomenon. However, a significant number of studies have found priming across modalities, although it is reduced, typically by half (Roediger & McDermott, 1993), relative to within-modality priming (Dijkstra, Schreuder, & Frauenfelder, 1989; Dodd, Oerlemans, & Robinson, 1988; Jackson & Morton, 1984; Jakimik, Cole, & Rudnicky, 1985; Kirsner & Smith, 1974; Monsell, 1985; Morton, 1979). This suggests that visual word processing can at least produce some degree of long-term change to phonological representations of the type that would be necessary to support a subtitling advantage.

The present experiments also examined repetition priming for unfamiliar words in order to gauge the effects of bimodal presentation on the implicit learning of new word forms. Nonword repetition priming within a single modality has been repeatedly demonstrated (e.g., Feustel, Shiffrin, & Salasoo, 1983; Kirsner & Smith, 1974; Monsell, 1985; Scarborough, Gerard, & Cortese, 1979; and for priming of nonwords in amnesics and normal subjects, see Haist, Musen, & Squire, 1991). However, it is generally less robust than word priming (Kirsner & Smith, 1974). Monsell (1985) suggests that this priming may reflect the early stages of new word form learning in some cases, and that these memory traces are "fragile" and therefore attenuate quickly and are easy to obliterate. Presumably, changing modalities, which generally results in markedly attenuated priming, even for known words, could obliterate the effect entirely. Monsell (1985) also suggests that repetition priming studies often discourage nonword priming by presenting large numbers of nonwords in a brief period of time. Thus, the absence of cross-modal nonword priming should not necessarily be taken as inevitable but possibly as a result of tasks that make too many demands on the processing, retention, and retrieval of unfamiliar word forms.

In summary, the present experiments differ from previous research on subtitling in that we were concerned with learning as opposed to comprehension. They also differ from previous research on repetition priming because we compared, not only within- and across-modality effects, but also the effect of combining modalities. A third way in which they differ is that we examined these effects in both native speakers of English and learners of English as a second language. This was because our aim was to evaluate the effectiveness of subtitling as a language learning tool.

EXPERIMENT 1

Method

Subjects. The subjects were 16 native speakers of English and 16 nonnative speakers of English. The native speakers were all university educated and ranged in age between 21 and 36. The nonnative speakers were recruited from a local language school and ranged in age from 18 to 24. All nonnatives were volunteers from the school's highest proficiency group, as determined by the school's proficiency test scores, making them roughly equivalent to subjects near the pass level in the Cambridge Proficiency examination. Twelve of the nonnatives were native speakers of Spanish and four were native speakers of Italian.

Materials. Two lists of 40 words were initially constructed from a dictionary search. One list was made up of "familiar" words and the other of "unfamiliar" words. Familiar words were ones judged likely to be very familiar to both the natives and nonnatives, and unfamiliar words were items predicted to be unfamiliar to both native and nonnative speakers. The items were a mixture of one and two syllables long. Selection was not done on the basis of frequencies because the subjects were not native speakers of English, so it seemed more reasonable to gauge how well individual words were known on the basis of feedback from pilot subjects of equivalent proficiency to the subjects in the experiment.

The two lists were then randomly combined in a single list, and a number of words of medium difficulty were inserted as fillers. The lists were shown to three native and three nonnative speakers of English, who acted as word raters. The nonnatives were all Ph.D. students at the University of Cambridge Research Centre for English and Applied Linguistics. The raters were asked to rate the words as *very familiar*, *less familiar*, and *completely unfamiliar*. Any familiar words that any rater judged as unfamiliar were discarded, as were any unfamiliar words rated as familiar. From this procedure, two sets of 20 words were selected, one familiar set, and one unfamiliar set (see Appendix A). An additional 12 familiar and unfamiliar words were used as practice items.

All of the words and nonwords were then recorded orally by the first author (in a Canadian accent) using a standard microphone and tape recorder and digitized using SoundEdit 16 on a Macintosh LC 575 computer. A text version of each word or nonword was also created in Geneva 12-point type. The text and sound files were incorporated into a program for running the experiment using SuperLab software.

In addition, a brief questionnaire was constructed to act as an interim task between Phases 1 and 2. The questions asked about the subjects' home country and native language, how they learned English, and whether they felt more proficient in reading or listening. Native speakers filled out a similar questionnaire asking about their foreign language learning experience.

Design. Equal numbers of familiar and unfamiliar words were assigned to each of four lists, and the items in each list were rotated around the four presentation conditions (sound, text, sound plus text, new) such that five items of each type appeared in each condition on each list. The native and nonnnative groups were each divided into four groups of four subjects. Each of the four groups of natives and four groups of nonnatives was randomly assigned one of the four lists, so that for each list there was a native and a nonnative group. Thus, subject group (native or nonnative) and list were between-subjects factors. Mode (sound, text, sound plus text, new) and word (familiar, unfamiliar) were within-subjects factors. In each list, five familiar and five unfamiliar words were assigned to each of four mode conditions. Ten additional items were used as fillers for Phase 2, eight were used as practice items, and two were used as buffers.

Procedure. Each subject was tested individually in a quiet room. Before beginning the practice phase, subjects were told that they would either see, hear, or simultaneously see and hear English words. They were asked to decide as quickly as possible whether they knew the meaning of the target word and to indicate their decision by pressing either the / key with the right hand for a familiar word or *z* key with the left hand for an unfamiliar word. Thus, if subjects did not know the meaning of a target in Phase 1, they would again respond "no" in Phase 2. In order to encourage subjects to answer honestly, they were also told that later in the experiment they would be asked to give the meanings for items for which they had answered "yes." They were encouraged to respond as quickly as possible, and not to spend time deciding whether an item might be an obscure English word.

For Phase 1 there was a 3-s blank interval between the offset of one item and the presentation of the next. The onsets of text and sound were synchronized, and response times were recorded from stimulus onset. The subject could respond to the stimulus at any point from onset to the end of the 3-s interval between trials. Text stimuli remained on the screen until the subject responded. Following Phase 1, the subjects filled out the questionnaire. This required approximately 2 min. The subjects were then told that Phase 2 would be identical to Phase 1, except that all stimuli would be presented as sound only.

At the end of Phase 2, subjects were given a 1-min break before the start of Phase 3. They were told that a list of words would be presented in the sound modality only. The task this time was to judge whether each item had been presented in Phase 1 of the experiment. This task was a surprise, as the subjects had not been asked to try to remember any items from Phase 1 or 2. Note that because the same items were used in Phases 2 and 3, new items in Phase 3 had in fact been presented as sound only in Phase 2. The subjects' task in Phase 3 was, therefore, to judge whether items had occurred in Phase 1, where no new items had occurred and the old items had occurred in different presentation conditions. The subjects were required to press the space bar if the item had been presented in Phase 1, or not press any key if the item had not been presented in Phase 1. Subjects were asked to respond as quickly and accurately as possible. Each item remained on the screen for 3 s in an effort to limit the decision time.

At the end of the experiment, subjects were shown the complete list of target words on paper and asked to put a check mark beside items for which they definitely knew the meaning. Any familiar words that the subjects reported not knowing were considered errors and later removed from the subjects' data. Similarly, any unfamiliar words that the subjects reported knowing the meaning of were also later removed from the data. Thus, at the end of the experiment, this procedure ensured that analyzed responses for familiar words were all true yes responses, the responses for unfamiliar words were all true no responses, and errors were response errors.

Results

For Phases 1 and 2, reaction times greater than 2.5 *SD* from the subject's mean for that condition were considered outliers and were removed. This resulted in a loss of nine outliers. Errors were also eliminated from the results of each of

Table 1. *Results from Phases 1 and 2 of Experiment 1*

	Familiar words				Unfamiliar words			
	S	T	S+T	New	S	T	S+T	New
				Phase 1				
Native								
RT	955	832	920		1103	908	955	
SD	81	55	87		119	107	84	
Error	.013	.013	0		.025	.013	.013	
Nonnative								
RT	1075	989	966		1195	1028	1019	
SD	75	97	56		103	79	64	
Error	.038	0	0		.062	.038	.013	
				Phase 2				
Native								
RT	924	1001	922	1008	1053	1063	1052	1068
SD	53	76	45	53	52	55	51	69
Error	.013	0	0	.025	.013	.013	0	.038
Priming	84	7	86		15	5	16	
Nonnative								
RT	1016	1059	1016	1077	1111	1121	1118	1128
SD	91	33	66	55	87	74	84	92
Error	.013	.025	.013	0	.05	.025	.025	.038
Priming	61	18	61		17	7	10	

Note: S, sound; T, text; S+T, sound plus text.

the three phases. The Phase 1 and 2 results for natives and nonnatives are shown in Table 1.

Phase 1. An overall analysis of variance, which included subject group and list as between-subjects factors, was used to analyze errors and reaction times. A 2 (Subject Group) × 4 (List) × 3 (Mode) × 2 (Word) analysis of variance (ANOVA) of error rates showed no significant main effects or interactions. This was apparently due to a ceiling effect for accuracy, because most subjects made no errors. The same ANOVA of reaction times showed a main effect of the subject group factor, indicating that natives were faster than nonnatives, $F(1, 24) = 97.60$, $p < .001$, and the word factor, $F(1, 24) = 50.40$, $p < .001$, indicating that reaction times for familiar words were faster than for unfamiliar words. The analysis also showed a main effect of mode, $F(2, 48) = 40.35$, $p < .001$. Neuman–Keuls (NK) comparisons showed that text and sound plus text were faster than sound ($p < .01$). There was also a significant interaction of mode and word, $F(2, 48) = 6.45$, $p < .01$. This appears to be because of the relatively slow responses to unfamiliar words in the sound condition.

Phase 2. Error rates for Phase 2 (see Table 1) were first submitted to a 2 (Subject) × 4 (List) × 4 (Mode) × 2 (Word) ANOVA. There were no significant

Table 2. *Results from Phase 3 of Experiment 1*

	Familiar words			Unfamiliar words		
	S	T	S+T	S	T	S+T
Native						
RT	1477	1552	1474	1512	1613	1485
SD	161	99	110	145	85	108
d′	1.237	.519	1.811	.431	.195	.621
Hit rate	.804	.545	.925	.579	.512	.656
False alarm rate	.346	.346	.346	.405	.405	.405
Nonnative						
RT	1478	1528	1453	1477	1589	1469
SD	129	164	168	166	140	176
d′	1.152	.455	1.545	.251	.172	.623
Hit rate	.781	.531	.875	.520	.484	.663
False alarm rate	.356	.356	.356	.420	.420	.420

Note: S, sound; T, text; S+T, sound plus text.

differences. As in Phase 1, accuracy was near ceiling for all subjects. Reaction times were then submitted to the same ANOVA. The analysis showed main effects of subject group, again indicating that natives were faster than nonnatives, $F(1, 24) = 64.01$, $p < .001$; mode, $F(3, 72) = 22.22$, $p < .001$; and word, $F(1, 24) = 119.03$, $p < .001$, the latter indicating that reaction times for familiar words were faster than for unfamiliar words. There was also an interaction of the mode and word factors, $F(3, 72) = 9.60$, $p < .001$. No other main effects or interactions were significant. The NK comparisons of the mode–word interaction showed that, for familiar words, the sound and sound plus text conditions had faster reaction times than the text condition and the new condition ($p < .01$), thus indicating a priming effect for familiar words in the sound plus text and sound conditions, but no priming effect in the text condition. No significant priming effect was found for unfamiliar words in any mode condition.

Phase 3. The reaction times for correct "old" responses, hit and false alarm rates, and *d′* values are shown in Table 2.[1] Native and nonnative reaction times for Phase 3 were submitted to a 2 (Subject) × 4 (List) × 3 (Mode) × 2 (Word) ANOVA. The results revealed a main effect of mode, $F(2, 48) = 11.48$, $p < .001$, and a main effect of word, $F(1, 24) = 12.36$, $p < .01$. No other main effects or interactions were significant. NK comparisons of the mode factor showed that sound and sound plus text were faster than text ($p < .01$). Thus the pattern of reaction times is similar to Phase 2: sound and sound plus text resulted in faster reaction times than text only.

The *d′* values for each presentation condition were also analyzed. A 2 (Subject) × 4 (List) × 3 (Mode) × 2 (Word) ANOVA of the *d′* values showed a main effect of mode, $F(1, 24) = 12.54$, $p < .001$, and a main effect of word,

$F(1, 24) = 42.84$, $p < .001$. NK comparisons of the mode factor showed that sound plus text was greater than sound ($p < .05$) and than text ($p < .01$). In addition, sound was greater than text ($p < .05$).

Discussion

The purpose of Experiment 1 was to test the claim that same-language subtitling improves listening comprehension (Garza, 1991). We tested subjects on three measures: Phase 1, reaction times to familiar and unfamiliar words, a measure of decoding speed; Phase 2, reaction times for old, previously presented and new items, a measure of implicit memory; and Phase 3, recognition memory for items presented in Phase 1, a measure of explicit memory.

In Phase 1, for both native and nonnative subjects, the fastest reaction times were found when text was available, that is, in the sound plus text and text conditions. This result suggests that both subject groups tended to rely on the visual stimulus to make their decisions. Faster decoding might be expected when textual support was provided simply because of the temporal constraints imposed by spoken word recognition. In order to decide that a stimulus was a word, subjects would have had to wait until they had heard enough information to be sure what word it was or, in the case of a nonword, they would have had to wait until such a point that the stimulus could not make any known word (see Marslen-Wilson, 1989, for a discussion of "uniqueness points" in relation to spoken word recognition processes). In contrast, when text was present, they could make a potentially more rapid judgment based on visual information. Faster decoding when words are presented as bimodal text and sound could be a factor in the improved comprehension found in subtitling studies (Borras & Lafayette, 1994; Danan, 1992). With more efficient decoding, more cognitive resources could be allocated to "higher level" processes (e.g., semantic processing), perhaps leading to better global comprehension (Perfetti, 1985).

Given that the Phase 1 reaction times strongly suggest that the subjects attended to, or even prioritized, the text when making lexical decisions in the sound plus text condition, one might have expected that this would have resulted in less priming of lexical decisions to spoken words in Phase 2 compared to the sound condition. However, in the case of words, the size of the priming effect was not significantly different in the two conditions. This may be taken to suggest that subjects were able to attend to, and fully process, both the text and sound, even though text was apparently strongly relied on for making lexical decisions in Phase 1. The combination of Phase 1 and 2 results lends support to the notion that language learners can benefit from bimodal input in terms of faster word recognition in Phase 1 without incurring any costs in terms of phonological information from the auditory modality retained in memory. This supports claims in the subtitling literature that subjects are not distracted from the audio input by subtitles (Vanderplank, 1988). However, the bimodal input failed to show any significant advantage over the sound condition, suggesting that the addition of text did not boost learning over that obtained with single-modality sound presentation.

The lack of priming for nonwords in any condition is consistent with studies showing that nonword priming is significantly attenuated or absent relative to words (Ellis, 1994). However, conclusions are difficult to draw from this result because, as Monsell (1985) suggests, repetition priming experiments typically discourage learning of new word forms because a large number of words and nonwords are randomly presented in a brief period of time. In addition, it is important to bear in mind that implicit memories are generally much slower to develop than explicit memories (Ellis, 1994). Our experiment seems to fit with the classic priming experiment that discourages learning in that a large number of familiar and unfamiliar items were randomly presented only once, and in rapid succession. The experimental design may therefore have been a key factor in the lack of nonword priming. This issue will be taken up in Experiment 2.

In Phase 3, native and nonnative subject groups both showed superior recognition memory scores in the sound plus text condition, compared to the sound and the text conditions. This result aligns well with Ellis' (1994) argument that explicit memory tasks permit access to elaborated features from presentation context, as demonstrated by the success of mnemonic devices such as the keyword method (Bird & Jacobs, 1999). Text would not by itself have prompted "deep" processing (Craik & Lockhart, 1972), but it may have provided a "richer" more elaborate and distinctive context that was beneficial to explicit recall (Craik & Tulving, 1975; Jacoby & Dallas, 1981). At the same time, sound showed superior recognition memory to text. This presumably reflects the fact that in the sound condition, the recognition memory target was identical to the stimulus heard in Phase 1.

EXPERIMENT 2

In this experiment we sought to establish whether there are any conditions under which repetition priming for unfamiliar words can be obtained using nonnative speakers of English and, if so, whether there are any differences between the sound and sound plus text Phase 1 presentation conditions. In the above discussion we suggested that one of the reasons why we failed to obtain repetition priming for unfamiliar words was that a single presentation may be insufficient to support the degree of implicit learning necessary to reveal a priming effect. In Experiment 2 we therefore presented nonwords three times during Phase 1.

However, if familiarity judgment is used as the Phase 2 task, then repeated presentation of nonwords in Phase 1 could actually interfere with responses to old as opposed to new items because of a feeling of familiarity associated with the old items. (See Balota & Chumbley, 1984, for a discussion of the role of familiarity in lexical decision.) Although this also applies to Experiment 1 (and could have provided another reason why no unfamiliar word priming effects were obtained), if in Experiment 2 each nonword is presented a number of times in Phase 1, then the problem is likely to be exacerbated. We therefore felt it useful to explore the effects of bimodal presentation using a different task, known as rhyme monitoring (Donnenwerth-Nolan, Tanenhaus, & Seidenberg, 1981; Seidenberg & Tanenhaus, 1979). In a rhyme monitoring task, subjects are presented with pairs of words (or nonwords) and asked to decide as quickly as possible whether the second item (the target) rhymes with the first item (the cue). For example, the subject might first be presented with a nonword cue such as *glemp*, followed by the target *fremp* and would have to decide as quickly as possible whether the target rhymes with the cue. If different cues are paired with the target each time it appears, enough novelty can be preserved in the decision for it to be used as a measure of repetition priming. The advantage of using this task in the present study is that any emerging familiarity with the target (e.g., *fremp*) should not conflict with a yes (it rhymes) decision. Thus, rhyme monitoring provides a means of testing nonword priming while avoiding conflicts between correct decisions and emerging word form representations.

Seidenberg and Tanenhaus (1979) found that word pairs that were orthographically similar (e.g., *stroke* and *joke*) were judged correctly as rhymes more quickly than rhyme pairs that were orthographically dissimilar (e.g., *soak* and *joke*), regardless of whether the rhymes and targets were presented aurally or visually. In terms of speed of responding, the optimal strategy for subjects would have been to make rhyme judgments solely on the basis of phonological information, yet they failed to do so and orthographic inconsistencies between rhyme and target items appeared to interfere with decisions even when no visual information was present in the cue or target. Seidenberg and Tanenhaus conclude from these results that orthographic information becomes available automatically when subjects are making auditory rhyme decisions. If orthographic information is activated automatically when subjects are processing auditory input, providing a simultaneous text modality (e.g., same-language subtitles) may not be as artificial or unusual a processing demand as one might imagine and it may be of assistance, particularly when subjects are unsure about the orthographic representation of a nonword. That is, providing textual representations may assist the subject in forming an orthographic representation that would normally be activated in the course of processing auditory information. In the rhyme monitoring task used here, all rhyming pairs were orthographically consistent.

Method

Subjects. Twenty-four advanced learners of English (as judged by the school's proficiency categories) were recruited from the same local language school as the subjects in Experiment 1. Twenty of the subjects were native speakers of Spanish, three were Italian, and one was Chinese. The subjects' ages ranged from 19 to 26 years. The subjects were from the same proficiency group as those in Experiment 1.

Materials. We constructed a set of 56 one- and two-syllable pronounceable nonwords to act as targets and another set of 10 to serve as practice items and buffers. The majority of the nonwords were two syllables long. For each of these nonwords we constructed two nonword rhymes to act as rhyming and nonrhyming cues (see Appendix B for targets and rhyme cues). Rhyme cues and their rhyming targets had identical numbers of syllables. Nonwords with ambiguous syllable boundaries (e.g., *mentast*, where *t* could belong to either syllable) were avoided to reduce the possibility of ambiguity in rhyme detection for speakers of some languages. (See Cutler, Mehler, Norris, & Segui, 1986, for evidence of language-specific segmentation strategies.) In addition, we constructed the nonwords so that they were feedforward consistent (Ziegler, Montant, & Jacobs, 1997); that is, there was (as far as possible) a single way of pronouncing the written nonwords, particularly the vowels. This was done to reduce pronunciation ambiguity in the text condition. The software and computer used for this experiment were identical to those used in Experiment 1. The only difference from Experiment 1 was that the cues and targets were spoken by a female with a British accent. This was done because the subjects were learners of English at a local language school and would probably be more familiar with a British accent.

Design. The basic design was a 4 (List) × 4 (Mode: Sound, Text, Sound + Text, New) × 2 (Rhyme/Nonrhyme) mixed factorial. The 24 subjects were randomly assigned to one of four lists (the between-subjects factor). All targets in each mode (except new) were presented four times over four cycles, once in each cycle. Cycles 1–3 will be referred to as Phase 1 (study phase) and Cycle 4 as Phase 2 (test phase). For each level of the rhyme factor, targets were paired with a rhyming cue in two cycles and a nonrhyming cue in the other two cycles. All of the rhyme cues were presented an equal number of times over the four cycles, but in half of the trials they were paired with rhyming targets and in the other half with nonrhyming targets. This was done so that, first, subjects would never be able to predict whether the following target would be a rhyme or a nonrhyme and, second, any given target had an equal chance of eliciting a yes (it rhymes) or no (it does not rhyme) response over the four cycles. The only difference between the two levels of the rhyme factor was the order in which pairs occurred: in the rhyme condition targets were paired with a rhyming cue in Cycles 2 and 4 and a nonrhyming cue in Cycles 1 and 3. In the nonrhyme condition targets were paired with a nonrhyming cue in Cycles 2 and 4 and a rhyming cue in Cycles 1 and 3. Thus, at Cycle 4 (i.e., Phase 2), items in the rhyme condition were paired with a rhyme (requiring a yes response) and items in the nonrhyme condition were paired with a nonrhyme (requiring a no decision).

The cues were presented as sound only in all four cycles. This was done to reduce the possibility of subjects developing a visual pattern matching strategy in the sound plus text and text conditions, as may have been the case if the cues had been presented visually. Therefore, only the targets were presented in presentation mode conditions during the first three cycles. In Cycle 4, all rhyme cues and targets were presented as sound only. Therefore, Cycle 4 corresponds to the Phase 2 presentation in Experiment 1 and the preceding cycles correspond to Phase 1.

The target items were semirandomly assigned either to the rhyme or nonrhyme conditions: after randomly assigning the items to conditions, the assignments were adjusted to ensure there were roughly equal proportions of one- and two-syllable items in each condition. The items were then rotated around the presentation conditions in four lists. However, the items were not rotated around the rhyme factor because this would have required eight lists and more subjects than were available for the experiment. Thus, while comparisons between the rhyme and nonrhyme conditions are confounded with items, the effect of presentation condition is not.

Cycles 1, 2, and 3 of Phase 1 each contained a block of 42 trials. In the first two cycles, the entire block of 42 paired rhyme/target trials was randomized. In Cycle 3, however, the items were partially randomized: target items were assigned to three separate groups of 14 items with equal proportions of items from each presentation condition in each block. Items were then randomized within their block of 14. This semirandom procedure allowed us to partially control the lag between each item's Cycle 3 and Cycle 4 presentations. This was important because trial lags appear to affect the strength of the repetition priming effect: short lags yield much greater priming than long lags, particularly in the case of nonwords (Monsell, 1985). Therefore, lag had to be held as constant as possible while still making predictive strategies very difficult by randomizing within the blocks of 14 trials.

In Cycle 4, or Phase 2, the randomization procedure from Cycle 3 was repeated along with 14 new nonwords, which had not been previously encountered. These were assigned randomly in almost equal proportions to each of the blocks of 14 (19 + 19 + 18 = 56 trials), and again, trials were randomized within each of these blocks. As in Experiment 1, priming effects were assessed for each condition by comparing old and new items. In Cycle 4, all rhymes and target cues were presented as sound only. This was the critical phase of the experiment, in which all previous presentation condition effects could be compared when subjects only heard the targets. Since Phase 1 involved a possible confound from uncontrolled item lag in Cycles 1 and 2, Phase 1 was treated only as a training phase, not as data for the analysis of the effects of presentation conditions on rhyme monitoring performance. Thus, we focused on performance data in Phase 2 (Cycle 4) and Phase 3.

As in Experiment 1, Phase 3 consisted of a recognition memory test. The same questionnaire was used as an intervening task as in Experiment 1. Targets were again randomized and presented as sound only, along with an equal number of new items that had not been presented in either Phase 1 or 2.

Procedure. Subjects were tested by the first author individually in a quiet room. Each subject was told that he or she would encounter pairs of pronounceable nonwords one pair at a time and were to decide upon encountering the second item whether it rhymed with the first item. In order to ensure that they were able to detect rhymes, the subjects were shown six pairs of nonwords printed on a piece of paper and asked to identify which pairs rhymed (none of these nonwords occurred in the actual experiment). All of the subjects performed the task perfectly.

The subjects were then told that nonword pairs would be presented by the computer. The first item of each pair would be presented auditorily, and the second would either be presented as sound only, as text only (on the computer screen), or simultaneously, as text and sound. The subjects were instructed that it was important to always watch the screen because sometimes the second item

Table 3. *Results from Phase 2 of Experiment 2*

	Rhyme				Nonrhyme			
	S	T	S+T	New	S	T	S+T	New
RT	952	902	902	1000	969	948	994	892
SD	261	171	179	201	248	31	195	193
Error	.083	.095	.03	.089	.071	.018	.042	.071
Priming	48	98	98		−77	−56	−102	

Note: S, sound; T, text; S+T, sound plus text.

would only be seen. The subjects were told to press the / key for a yes response or the z key for a no response. The subjects were then given a practice phase consisting of six trials (two from each mode condition, one paired with a rhyme cue and one with a nonrhyme cue). The subjects were then given an opportunity to ask questions to make sure they understood the task, and were again reminded to watch the screen throughout and to respond as quickly and accurately as possible.

Four buffer trials were used at the beginning of each phase: two rhyming pairs and two nonrhyming pairs, consisting of one pair in each of the text and sound conditions and two in the sound plus text condition. There were no breaks between cycles. In each trial, the cue item was presented auditorily until its offset. The target began after a 1.5-s interstimulus delay. In the sound plus text condition the auditory and visual stimulus onsets were synchronized, and for all conditions the reaction times were measured from the onset of the target stimulus until the subject made a response. Whenever text was present, it remained on the screen until the offset of the sound stimulus.

At the end of the rhyme judgment task (Phases 1 and 2), the subjects were asked to fill out a questionnaire (identical to that used in Experiment 1) in order to give them a brief break of about 3 min.

Phase 3 was a recognition memory task identical in design to that in Experiment 1. All 56 targets and an equal number of new nonword targets were presented, all as sound only, and subjects were asked to decide as quickly as possible whether they remembered having encountered each item (regardless of whether it had been seen and/or heard) in the rhyme judgment task. The intertrial interval was 1.5 s. In addition to accuracy, reaction times were also recorded from stimulus onset.

Results

Reaction times exceeding 2.5 standard deviations from the subject's mean for that condition were considered outliers and were excluded from the data analysis. This resulted in a loss of 13 reaction times over the four cycles of Phases 1 and 2 and 3 reaction times from Phase 3. All errors from both phases were also excluded. Table 3 provides a summary of the results from Phase 2 (i.e., Cycle 4).

Phase 2. The reaction time data for the Phase 2 rhyme and nonrhyme conditions were submitted to an ANOVA in which list (1–4) was a between-subjects factor and mode (sound, text, sound plus text, new) and rhyme (rhyme, nonrhyme) were within-subjects factors. There was a significant interaction between mode and rhyme, $F(3, 60) = 5.31$, $p < .01$. No other main effects or interactions were significant.

The reaction time data for the Phase 2 rhyme and nonrhyme conditions were then submitted to separate analyses of variance. In the rhyme condition, there was a main effect of mode, $F(3, 60) = 4.83$, $p < .01$. NK comparisons showed significantly faster reaction times in the text and sound plus text conditions compared to new ($p < .05$), indicating priming effects. In Phase 2, therefore, nonword repetition priming was obtained in the text and sound plus text conditions but not in the sound condition. There was no significant difference between the sound plus text and text conditions or between sound and sound plus text or text. This pattern of results is at odds with Experiment 1 and most previous studies, which have typically shown twice as much priming within modalities compared with cross-modal priming (Roediger & McDermott, 1993).

The ANOVA of the nonrhyme condition showed a rather different pattern. There was again a main effect of mode, $F(3, 60) = 2.97$, $p < .05$. However, NK comparisons showed that reaction times were significantly faster for new than for sound plus text ($p < .05$). No other comparisons were significant. This result indicates an interference effect for sound plus text relative to new items and no significant effect of any kind in the sound and text conditions.

The error rates were also analyzed. A 4 (List) × 4 (Mode) × 2 (Rhyme) ANOVA showed a main effect of Mode, $F(3, 60) = 4.47$, $p < .01$. NK comparisons showed fewer errors for sound plus text compared with sound and new ($p < .05$), but no significant difference compared with the text condition.

Separate analyses of variance were then carried out on the error rates for the rhyme and nonrhyme conditions, particularly to explore the noticeably unbalanced error rates in the text condition over the rhyme and nonrhyme conditions.[2] The results showed a marginally significant main effect of mode in the rhyme condition, $F(3, 60) = 2.74$, $p = .051$, with sound plus text the lowest error rate. In the nonrhyme condition, there was a main effect of mode, $F(3, 60) = 3.68$, $p < .05$. NK comparisons showed significantly fewer errors for text compared with sound and new. No other comparisons were significant. The text condition therefore had a strong beneficial effect on error reduction in the nonrhyme condition but no effect in the rhyme condition (at least relative to new items). By contrast, the sound plus text condition had a fairly even beneficial effect on error reduction across the rhyme and nonrhyme conditions, as demonstrated by the ANOVA when scores were collapsed over the rhyme/nonrhyme conditions.

Phase 3. The mean reaction time for correct old responses, hits, false alarm rates, and d' values are shown in Table 4. An analysis of variance of reaction times with respect to the three mode conditions from Phase 1 showed no significant differences. Thus, the speed of recognition memory responses was not differentially affected by the Phase 1 presentation mode conditions.

Table 4. *Results from Phase 3 of Experiment 2*

	Rhyme			Nonrhyme		
	S	T	S+T	S	T	S+T
RT	1108	1059	1096	1080	1101	1135
SD	139	175	150	211	171	222
d′	.953	.523	1.825	.913	.596	.977
Hit rate	.757	.606	.947	.746	.635	.768
False alarm rate	.398	.398	.398	.406	.406	.406

Note: S, sound; T, text; S+T, sound plus text.

The *d′* values were first analyzed with a 4 (List) × 3 (Mode) × 2 (Rhyme) ANOVA. There was a significant interaction between the mode and rhyme factors, $F(2, 40) = 6.36$, $p < .01$. NK comparisons showed that in the rhyme condition, *d′* scores were significantly higher in the sound plus text condition relative to sound and text scores ($p < .01$). In addition, sound was significantly higher than text ($p < .05$), and sound plus text/rhyme was significantly higher than sound plus text/nonrhyme ($p < .01$). There were no significant differences in the nonrhyme condition.

Discussion

Experiment 2 used a rhyme monitoring task to explore the possibility that the learning of spoken nonwords might be observed if subjects were presented with targets three times prior to test. Some learning was in fact observed and was apparently affected by presentation modality and the context of the learning task. The patterns of implicit and explicit memory performance are discussed below.

Considering first the Phase 2 rhyme conditions, subjects showed implicit memory facilitation for nonwords in two of the three mode conditions, text and sound plus text. This result demonstrates that under certain experimental conditions, it is possible to observe the kind of implicit performance facilitation normally associated with words in lexical decision tasks (as in Experiment 1). Monsell (1985, p. 166) suggested that “under some conditions a person may, when presented with a nonword, learn it as such. Learning might consist of forming a (possibly fragile) new lexical unit, as must frequently happen when we encounter new words in text or discourse.” The present experiment presumably provided a more learning-conducive context by exposing subjects to the target items three times prior to test.

As predicted by researchers who suggest that bimodal presentations can bolster later spoken word recognition (e.g., Garza, 1991), Experiment 2 demonstrated significant facilitation for nonwords that had earlier been encountered as sound and text simultaneously, but not for nonwords that had been encountered as sound alone. This is in some ways an odd result. A number of researchers have argued that the magnitude of priming effects will depend on the degree to which the study "episode," which includes targets, context, and task-specific processing, is the same as the test episode (e.g., Jacoby, 1983); and most cross-modal priming studies have generally found the strongest priming when study and test modality are identical (Roediger & McDermott, 1993). Thus, we might have expected the sound condition to have had the strongest priming effects because the training encounters with the targets in Cycles 1–3 were identical to the critical encounter in Cycle 4.

A plausible explanation of the priming in the sound plus text condition is that over the first three cycles the text acted as a comprehensible information source, allowing the subjects to begin to develop a superior memory trace for spoken nonwords because they were more certain of what they were hearing (see, for example, Frost, Repp, & Katz, 1988). This points to the kind of "unconscious integration of the senses" proposed by Vanderplank (1988). The fact that error rates in the text and sound plus text conditions were significantly reduced relative to the sound condition lends support to this interpretation. The addition of text may have allowed subjects to resolve phonological ambiguities in some of the spoken targets. The lack of priming in the sound condition might therefore have been because no additional information was available to aid the resolution of ambiguity. With nonnative speakers of English, it seems reasonable to assume that if subjects are unable to correctly identify rhyming pairs of individual words in the first few exposures, they have probably misheard or incorrectly decoded the auditory signal. Learning then becomes difficult and repeated exposure to targets may only repeat the process of being unsure of what is being heard. Presumably, uncertainty can result in slowed reaction times and more errors. The fact that Phase 2 error rates for sound were apparently unaffected by Phase 1 study episodes, relative to new items, again suggests that if word pairs were not heard and judged correctly the first time, they were never heard correctly.

We suspect that the priming effect in the text condition may have been a reflection of the importance of phonology in the rhyme monitoring of the task. When presented with pairs of written words and asked to make rhyme judgments, it seems highly likely that subjects would have been encouraged to convert orthographic information to phonological information. (See Frost, 1998, for a discussion of how tasks can influence the amount of phonological information activated during orthographic processing.) MacLeod and Masson (2000) found equivalent priming effects for "read" (i.e., simply reading visually presented words aloud) and "generated" items (i.e., produced by subjects in response to a semantic cue) on a speeded word-reading test. The authors argued that generated information can, under certain circumstances, be equivalent to perceptual information, and therefore produce priming effects that are equivalent to the priming observed when items have been read during study. Similarly, Monsell (1985) found that auditory lexical decisions were primed by a visual study condition. Monsell argued that under some circumstances, subjects may automatically convert print to phonological information, and this information can be used to facilitate auditory recognition as if subjects had heard the targets during study. In the present experiment, a rhyme judgment task would presumably be even more likely to encourage orthography–phonology conversion because the task required subjects to make decisions about phonological information.

Thus, by Phase 2 the priming effect in the text condition was probably the result of subjects having generated phonological representations for themselves. Thus, to some extent the text condition may have resembled a bimodal input condition in that subjects relied on orthographic and (generated) phonological information.[3]

Turning to the nonrhyme conditions, there was a general tendency for old targets to produce longer responses than new ones, although this effect was only significant for the sound plus text condition. It should be noted that in this condition the nonrhyming targets in Phase 2 had been encountered once with nonrhyming cues in Phase 1 but twice with rhyming cues. Morris, Bransford, and Franks (1977) found that memory for words that had been studied in a yes rhyme context yielded facilitation, whereas the no rhyme context showed no facilitation. Morris et al. argue that "congruence" in encoding context may be a significant factor in the strength of memory traces. That is, yes responses may yield qualitatively better memory traces than no responses because the cue and target fit together in the study context. By Cycle 4 the cue items may have encouraged subjects to expect a rhyming target and bias them to respond yes when the correct response was no. However, the fact that significant interference was found only in the sound plus text condition may be regarded as further evidence for more robust memory traces. That is to say, interference effects can be interpreted as improved learning when the required response runs counter to the biases created by the study condition (Jakimik et al., 1985).

The markedly different error rates for text over the rhyme and nonrhyme conditions suggest that, whereas generated phonological representations were adequate for accurately rejecting nonrhyme pairs, they were not a close enough match to Phase 2 spoken targets to allow subjects to reliably make correct rhyme judgments. It may be that correctly rejecting nonrhyme pairs requires a less fine-grained analysis of cues and targets. That is to say, internally generated phonological information does not need to be precisely matched with target phonology to assist in the correct rejection of nonrhymes, whereas a correct yes for rhyming pairs probably required a detailed match between all target rime phonemes and memory for the rime cluster from Phase 1. Presumably, the best kind of information for detailed matching processes would be a phonological memory trace derived from having correctly heard the targets during study. The fact that error rates were no better in the sound condition is an indication that sound alone did not allow subjects to correctly hear all targets. Apparently, only in the sound plus text condition was enough information available to allow subjects to develop accurate memories for the precise phonological information required to improve accuracy in yes rhyme judgments.

The pattern of results for Phase 3 was very much the same as that found in Experiment 1: the sound plus text condition yielded better recognition memory performance than the other two mode conditions. However, an effect of mode was only obtained in the rhyme condition. These results partially accord with Morris et al. (1977), who found improved recognition memory only for targets that had been studied in "congruent" rhyme contexts, that is, when targets were studied in a yes rhyming context (see also Craik & Tulving, 1975). Morris et al. did not elaborate on the precise role of congruent context at the time of encoding. In the present context, we simply take note of the possibility that recognition memory can be affected differently by yes and no study contexts.

General discussion

Two experiments were concerned with testing the effects of different presentation modes on implicit and explicit memory. The results of the experiments are discussed in terms of these two aspects of memory and their importance for foreign language vocabulary acquisition.

With regard to implicit memory, the two experiments produced somewhat different results. In Experiment 1, repetition priming from auditory words was not increased by the addition of text; whereas in Experiment 2, bimodal input boosted the implicit learning of nonwords compared to sound only. It appears that bimodal presentation beneficially affected implicit memory only when new phonological forms needed to be encoded. These results suggest that the effect of text is limited to cases in which the phonological form of the input can not be reliably established on the basis of the sound alone. For the familiar words in Experiment 1, the target words in Phase 1 could be reliably recognized, even in the absence of textual support, as the low Phase 1 error rates show. Therefore, even if phonological information had been generated from the text, it would not have provided any information over and above that available from the sound input. In Experiment 2, phonological information generated from the text was clearly being used to help establish phonological representations of the nonwords, presumably because the subjects were less able to form stable representations of those forms from sound input alone. Indeed, the priming effect on reaction times from text alone shows that generated phonology can have as large an impact on subsequent auditory processing as bimodal presentation. What appears to be critical is not the origin of the phonological information but whether it is sufficiently precise to contact a preexisting representation, be that of either a familiar word or a nonword that may only have been experienced one or two times previously. The provision of textual support increases the probability that this will be the case for nonwords, whereas for the familiar words in Experiment 1 the auditory input alone was sufficient.

The present results suggest, therefore, that phonological information derived from both text and sound contributes to improvements in the processing of spoken words, as subtitling researchers have speculated. Although the issue of orthographic and phonological interactions has been much debated in the priming literature (Monsell, 1985; Morton, 1979; Roediger & McDermott, 1993), evidence is now available to support the idea that the cognitive systems dealing with auditory and visual word recognition are highly interactive and fully interconnected (Masson, 1995; Van Orden, Pennington, & Stone, 1990). Indeed, experimental evidence has led to models that no longer regard orthographic and phonological processes as distinctly dual routes (Coltheart, 1978). For example, Masson's (1995, p. 6) distributed memory model emphasizes that orthographic and phonological information "converge on a single set of processing units so that their influence is combined." Admittedly, interactive activation models such as Masson's generally stem from studies of phonological effects on visual word recognition. (See also McClelland & Rumelhart, 1981, and Grainger & Ferrand, 1996, for slightly different versions of the same basic model.) However, Seidenberg and Tanenhaus (1979) demonstrated similar influences in the reverse direction (orthographic influences on auditory rhyme detection). More recently, Schneider and Healy (1993) demonstrated that phoneme detection can be influenced by spelling. Experiment 2 of the present study suggests an influence of orthographic information on spoken nonword recognition. We therefore see reason to suspect that orthographic and phonological information may have reciprocal influences.

It could be argued, however, that the present results do not show that information from text and sound were actually combined during processing in the Phase 1 tasks of Experiments 1 and 2. In both experiments the repetition priming effects on reaction time were never greater for bimodal than single modality presentation in Phase 1; that is, relative to sound only in Experiment 1 or text only in Experiment 2. Therefore, only one modality may have been responsible for the priming effects (sound in Experiment 1 and text in Experiment 2), and information from different modalities may not actually have combined during processing. Strictly speaking, then, it can only be argued that phonological information activated at encoding has an impact on the subsequent processing of spoken forms, regardless of whether it is derived from sound, text, or possibly both. Obviously, for this to happen, the information derived from text and sound must ultimately converge on a common representation, but this does not necessarily mean that the two information sources interacted and had reciprocal influences on each other during processing. Having said this, it is possible that an advantage of bimodal presentation over text or sound alone would be more pronounced when the sound and text input are both ambiguous when taken in isolation, for example, when the sound is degraded and words with ambiguous (feedforward inconsistent) spelling patterns are used. Each modality might then be expected to compensate for deficiencies in the other.

With regard to explicit memory, the results of the two experiments demonstrated that explicit memory performance in the form of a recognition memory task was aided by earlier bimodal input. It is well known that more "elaborated" stimulus contexts seem to aid accuracy in later conscious recall (see Jacoby, 1983). Experiment 1, Phase 3, demonstrated this effect for words; Experiment 2, Phase 3, demonstrated it for nonwords. We can assume that the bimodal input provided a more elaborate and useful stimulus for subjects to use when deciding whether a particular auditorily presented item came earlier in the experiment. Conscious recall is what has typically been tested in the subtitling literature. What is new in the present study is, first, that recognition memory of auditory information was improved, and, second, that there was an effect of an enriched memory trace at a level below that of semantic information, independent of any semantic context. By removing the semantic context that was available in the subtitling studies, the present experiments show more clearly that providing subjects with text and sound versions of known and unknown words can facilitate recognition memory relative to sound alone.

For subtitling research and practical questions about teaching listening comprehension, the important outcome of the present experiments is that the bimodal condition created no apparent interference with auditory processing and learning and led to improved implicit learning of novel word forms when the experiment allowed targets to be recycled three times prior to the test. Fears that dividing attention between modalities might interfere with listening comprehension were not borne out in the present study. Therefore, the present results demonstrate that bimodal inputs can be attended to and used to bolster both the implicit and explicit aspects of vocabulary learning.

Assuming that some of the effects observed in the present experiments transfer to the processing of words in sentence contexts, we have good reason to suggest that learners of a foreign language may in some ways benefit from same-language subtitling and other materials that allow learners to read and hear simultaneously. Many listening materials texts encourage subjects to use listening strategies in the context of authentic listening materials, in the apparent belief that the listening comprehension ability of native speakers is largely dependent on the processing of phonological and phonetic information along with semantic context clues (e.g., Mendelsohn, 1994). The evidence presented here and in other repetition priming studies suggests that orthographic information can, under certain circumstances, have a significant facilitatory impact on the long-term implicit, and explicit, learning of spoken word forms. In light of this evidence, materials that encourage a more integrated visual and auditory learning context, such as same-language subtitling, might be attractive alternatives to traditional listening materials. Teachers should note that most DVDs now include same-language subtitling as an optional viewing feature. The subtitling is intended for the hearing impaired; but as the present study suggests, this may also be of benefit to second language learning.

APPENDIX A

The familiar and unfamiliar critical target words used in Experiment 1

Familiar words			
Heater	Curtain	Apple	Flower
Business	Pencil	Success	Market
Music	Money	Table	Teacher
Printer	Story	Paper	Ceiling
Question	Party	Winter	Police
Unfamiliar words			
Clinid	Limner	Merkin	Hemin
Pavis	Costrel	Scolex	Palsen
Tomtate	Caroon	Chaptrel	Pongid
Frustum	Mackle	Corban	Mazard
Cooter	Saurel	Vection	Flavone

APPENDIX B

The targets and rhyme cues used in Experiment 2

Target	Rhyme cues	Target	Rhyme cues
1. Segtem	Regtem, megtem	29. Strimp	Zimp, climp
2. Mester	Nester, sester	30. Porping	Glorping, sorping
3. Astic	Pastik, bastic	31. Ramlin	Dramlin, lamlin
4. Bick	Fick, twick	32. Lince	Fince, glince
5. Mence	Lence, quence	33. Flanster	Nanster, panster
6. Twate	Zate, glate	34. Vertim	Nurtim, dertim
7. Yoaden	Moden, droden	35. Bervit	Dervit, hurvit
8. Tibber	Libber, nibber	36. Mordom	Hordom, tordom
9. Diller	Jiller, twiller	37. Venlin	Henlin, senlin
10. Wingle	Tringle, gringle	38. Bintor	Lintor, jintor
11. Drenton	Wenton, senton	39. Stog	Pog, crog
12. Sompest	Lompest, dompest	40. Exter	Wekster, pekster
13. Narject	Parject, garject	41. Flant	Sant, trant
14. Framp	Gamp, sqwamp	42. Steek	Pleek, neak
15. Kalloy	Galloy, estroy	43. Crad	Stad, shad
16. Demper	Hemper, femper	44. Dorcam	Slorkam, norkam
17. Yith	Bith, spith	45. Bistan	Histan, wistan
18. Hepping	Jepping, nepping	46. Merp	Glurp, surp
19. Actle	Nacktle, spacktle	47. Trindest	Sindest, quindest
20. Tillan	Fillan, nillan	48. Vorpid	Norpid, horpid
21. Glanful	Sanful, tranful	49. Garcust	Larcust, barcust
22. Wilk	Tilk, zilk	50. Arvon	Flarvon, sarvon
23. Mentast	Fentast, sentast	51. Striggle	Liggle, fliggle
24. Shastin	Lin, kentin	52. Venlem	Renlem, krenlem
25. Trongle	Slongle, wongle	53. Lorp	Sorp, glorp
26. Burrent	Durrent, turrent	54. Flanzen	Tanzen, banzen
27. Barnage	Garnage, farnage	55. Scrillit	Dillit, shillit
28. Crund	Twund, mund	56. Sippy	Glippy, vippy

NOTES

1. The false alarm rate is the rate of responding *old* to new items, and as these had not occurred during Phase 1, the false alarm rates in Table 2 are the same across all presentation conditions.
2. Note that, whereas the rhyme factor is confounded with items because of the experiment's design, only in the text condition were error rates radically different. This suggests that the difference in error rates was not related to item differences because items were rotated around the mode conditions.
3. Given that the priming effect was identical in the text and sound plus text conditions, it could be argued that in both cases priming was entirely due to phonological information generated from the text. This point is addressed in the General Discussion, but note that the pattern of error rates suggests that bimodal presentation did, in fact, confer an advantage over text alone (see below in text).

REFERENCES

Allport, D. A., & Funnell, E. (1981). Components of the mental lexicon. *Philosophical Transactions of the Royal Society of London B295,* 397–410.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Bird, S. A., & Jacobs, G. M. (1999). An examination of the keyword method: How effective is it for native speakers of Chinese learning English? *Asian Journal of English Language Teaching, 9,* 75–97.

Borras, I., & Lafayette, R. C. (1994). Effects of multimedia courseware subtitling on the speaking performance of college students of French. *The Modern Language Journal, 78,* 61–75.

Burnage, G. (1997). *Integrating digital video into language learning software.* Available at the University of Cambridge Faculty of Modern and Medieval Languages website.

Chung, J. (1996). The effects of using advance organizers and captions to introduce video in the foreign language classroom. *TESL Canada Journal, 14,* 61–65.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed), *Strategies of information processing* (pp. 151–216). New York: Academic Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

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

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning, 42,* 497–527.

Davis, R. S. (1998). Captioned video: Making it work for you. *The Internet TESL Journal, 4*(3). Retrieved February 6, 1998, from http://iteslj.org/techniques/davis-captionedvideo/

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

Dodd, B., Oerlemans, M., & Robinson, R. (1988). Cross-modal effects in repetition priming: A comparison of lipread, graphic and heard stimuli. *Visible Language, 22,* 58–77.

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.

Ellis, A. W. (1982). Modality-specific repetition priming of auditory word recognition. *Current Psychological Research, 2,* 123–128.

Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages.* London, Academic Press.

Feustel, T. C., Shiffrin, R. N., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General, 112,* 309–346.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition. *Psychological Bulletin, 123,* 71–99.

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.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24,* 239–250.

Gipson, P. (1984). *A study of the long-term priming of auditory word recognition.* Unpublished doctoral dissertation, University of Cambridge.

Grainger, G., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language, 35,* 623–647.

Haist, F., Musen, G., & Squire, L. R. (1991). Intact priming of words and nonwords in amnesia. *Psychobiology, 19,* 275–285.

Holobow, N. E., Lambert, W. E., & Sayegh, L. (1984). Pairing script and dialogue: Combinations that show promise for second or foreign language learning. *Language Learning, 34(4),* 59–74.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory and Cognition, 12,* 568–574.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behaviour, 22,* 485–508.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3,* 306–340.

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language, 24,* 165–178.

Jung, U. (1990). The challenge of broadcast videotex to applied linguistics. *IRAL, 28,* 201–219.

Kirsner, K., & Dunn, J. C. (1985). The perceptual record: A common factor in repetition priming and attribute retention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance* (Vol. XI, pp. 547–565). Hillsdale, NJ: Erlbaum.

Kirsner, K., Milech, D., & Standen, P. (1983). Common and modality-specific processes in the mental lexicon. *Memory and Cognition, 11,* 621–630.

Kirsner, K., & Smith, M. C. (1974). Modality effects in word identification. *Memory and Cognition, 2,* 637–640.

Lambert, W. E., Boehler, I., & Sidoti, J. (1981). Choosing the language of subtitles and spoken dialogues for media presentations: Implications for second language education. *Applied Psycholinguistics, 2,* 133–148.

MacLeod, C. M., & Masson, M. E. J. (2000). Repetition priming in speeded word reading: Contributions of perceptual and conceptual processing episodes. *Journal of Memory and Language, 42,* 208–228.

Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In M. Marslen-Wilson (Ed), *Lexical representation and process* (pp. 25–74). Cambridge, MA: MIT Press.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 3–23.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1: An account of the basic findings. *Psychological Review, 88,* 365–407.

Mendelsohn, D. (1994). *Learning to listen.* San Diego, CA: Dominie Press.

Monsell, S. (1985). Repetition and the lexicon. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 147–195). London: Erlbaum.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Level of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behaviour, 16,* 519–533.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstad, & H. Bouma, *Processing models of visible language* (pp. 259–268). New York: Plenum Press.

Perfetti, C. A. (1985). *Reading ability.* Oxford: Oxford University Press.

Price, K. (1983). Closed caption T. V.: An untapped resource. *MATSOL Newsletter, 12,* 1–8.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45,* 283–331.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Eds.), *Handbook of neuropsychology,* (Vol. 8, pp. 63–130). Amsterdam: Elsevier.

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory and Cognition, 7,* 3–12.

Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience, 4,* 244–256.

Schneider, V. I., & Healy, A. F. (1993). Detecting phonemes and letters in text: Interactions between different types and levels of processes. *Memory and Cognition, 21,* 739–751.

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

Vanderplank, R. (1988). The value of teletext subtitles in language learning. *ELT Journal, 42,* 272–281.

Vanderplank, R. (1990). Paying attention to the words: practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) subtitles. *System, 18,* 221–234.

Vanderplank, R. (1993). A very verbal medium: Language learning through closed captions. *TESOL Journal, 3,* 10–14.

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.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37,* 533–554.