651

JOURNAL OF MEMORY AND LANGUAGE 27, 741–755 (1988)

Can Speech Perception Be Influenced by Simultaneous Presentation of Print?

RAM FROST, BRUNO H. REPP, AND LEONARD KATZ

Haskins Laboratories, New Haven, Connecticut

When a spoken word is masked by noise having the same amplitude envelope, subjects report that they hear the word much more clearly if they see its printed version at the same time. Using signal detection methodology, we investigated whether this subjective impression reflects a change in perceptual sensitivity or in bias. In Experiment 1, speech-plus-noise and noise-only trials were accompanied by matching print, nonmatching (but structurally similar) print, or a neutral visual stimulus. The results revealed a strong bias effect: The matching visual input apparently made the amplitude-modulated masking noise sound more speechlike, but it did not improve the detectability of the speech. However, reaction times for correct detections were reliably shorter in the matching condition, suggesting perhaps subliminal facilitation. The bias and reaction time effects were much smaller when nonwords were substituted for the words, and they were absent when white noise was employed as the masking sound. Thus it seems that subjects automatically detect correspondences between speech amplitude envelopes and printed stimuli, and they do this more efficiently when the printed stimuli are real words. This supports the hypothesis, much discussed in the reading literature, that printed words are immediately translated into an internal representation having speechlike characteristics.

In the process of recognizing spoken words the listener must generate from the acoustic signal an internal representation that can make contact with the entries in the mental lexicon. A question of great importance for contemporary theories of speech perception is whether or not the generation of that representation is independent of lexical processes. One possibility is that the perceptual analysis of the speech input is completed before any contact with the mental lexicon occurs. Alternatively, some or all stages of the perceptual analysis may be interactively influenced by lexical processes that have been set in motion by partial information, prior context, or expectations. (See Frauenfelder & Tyler, 1987, for a review.)

This work was supported in part by National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. Correspondence concerning this article and reprint requests should be addressed to Ram Frost, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695.

Researchers concerned with auditory word perception generally take it for granted that the representations of words in the mental lexicon are phonologic in nature. (A notable exception is Klatt, 1980.) Investigators of visual word perception, too, often assume that phonological representations are accessed, although sometimes they postulate the existence of a separate visual–orthographic lexicon. While there are results indicating rapid visual recognition of written words prior to phonological analysis in certain tasks, there is much evidence that reading involves a phonological lexicon at some stage. (See McCusker, Hillinger, & Bias, 1981, for a review.) Theoretical parsimony dictates that this lexicon be the same as the one accessed in auditory word recognition. If so, then the process of speech perception might be penetrable to visual influences (as well as the reverse): If earlier stages of auditory word perception can be affected by lexical processes, and if those same lexical processes can be activated in parallel by a

0749-596X/88 $3.00

visual presentation of print, then perception of words in the auditory modality could be influenced by words presented in the visual modality.

Evidence for lexical top-down effects within the auditory modality has been obtained in tasks involving phoneme restoration (Warren, 1970; Samuel, 1981), detection of mispronunciations (Cole & Jakimik, 1980), and shadowing (Marslen-Wilson & Welsh, 1978). It is not known, however, whether similar effects on speech perception can be elicited by visual input. That visual and auditory speech information can interact at a rather early level in perception has been demonstrated by McGurk and MacDonald (1976): The visual presentation of articulatory gestures (a speaker's face) can affect subjects' perception of speech segments, even when the auditory input is unambiguous. However, because speech gestures are fundamental correlates of phonetic categories, their effect on speech perception takes place even before phonetic categorization, and certainly before lexical access (see Summerfield, 1987). The mapping of print into speech is far less direct and must be mediated by a lexical phonological level. Can a simultaneous presentation of printed words nevertheless influence the perception of speech?

A recent study by Frost and Katz (submitted) suggests that it might. These authors presented printed and spoken words simultaneously and asked subjects to judge whether the words were the same or different. The experiment included a condition in which the speech was degraded severely by added signal-correlated noise (a broadband noise with the same amplitude envelope as the stimulus). Nevertheless, the subjects found the task fairly easy; the average error rate was only 10 percent. In a subsequent pilot study, the same authors presented the subjects simultaneously with both degraded speech *and* degraded print. Here, subjects' performance was close to chance. The subjects' phenomenological description was that they often could not hear any speech at all in the auditory input, whereas previously, when clear print matching a degraded auditory word was presented simultaneously, they reported no difficulty in identifying the degraded word. Thus, it seemed as if the presence of the printed word enabled the subjects to separate the speech from the noise, and hence to perceive it much more clearly. There is another possibility, however: Subjects' introspections may have reflected merely an illusion caused by the correspondence between the print and the amplitude envelope of the masking noise, which was identical with that of the speech. That is, subjects might have thought they heard speech even if the masking noise alone had been presented accompanied by "matching" print; however, no such trials were included.

The following experiments were conducted to determine whether simultaneous presentation of the printed word can truly facilitate the detection of a speech signal in noise. A positive answer to this question would provide strong support for the hypothesis that visual and auditory word perceptions are functionally interdependent. On the other hand, even the finding of a pure "response bias" elicited by the correspondence between the print and the noise amplitude envelope would be interesting, as it, too, represents an effect of print on speech perception, though of a different kind.

Experiment 1

In Experiment 1, subjects were presented with either speech plus noise or with noise alone, and their task was to detect the presence of the speech signal. In conjunction with the auditory presentation, the subjects saw on a computer screen a matching word, a nonmatching word, or a neutral stimulus (XXXXX). The purpose of the experiment was, first, to confirm our pilot observations that simultaneous presentation

of matching print makes subjects hear speech in the noise and, second, to examine whether that effect reflects an increase in sensitivity to speech actually present or a bias to interpret the masking noise as speech. The signal detection paradigm we used enabled us to compute independent indices of sensitivity and bias.

It is important to keep in mind that even on noise-only trials there could be a (partial) match between the visual and the auditory stimuli, since each spoken word had its individual envelope-matched masking noise. Another way of framing the question, therefore, was to ask whether matching print would enhance the detectability of the *spectral* features of speech hidden in the noise, or whether it would have an effect on the subjects' responses because the auditory *amplitude envelope* corresponds to that of the spoken form of the printed word. These two effects are not mutually exclusive and may operate simultaneously.

Methods

Subjects. Thirty-six undergraduate students, all native speakers of English, participated in the experiment for payment.

Stimulus preparation. The stimuli were generated from 24 regular, disyllabic English words that had a stop consonant as their initial phoneme. All words were stressed on the first syllable. The number of phonemes in each word ranged from four to six, and the word frequencies, according to Kucera and Francis (1967), ranged from 0 to 438, with a median of 60. The words were spoken by a female speaker in an acoustically shielded booth and were recorded on an Otari MX5050 tape recorder. They were then digitized at a 20-kHz sampling rate. From each digitized word, we created a noise stimulus with the same amplitude envelope by randomly reversing the polarity of individual samples with a probability of 0.5 (Schroeder, 1968). Such signal-correlated noise retains a certain speechlike quality, even though its spectrum is flat and it cannot be identified as a particular utterance unless the choices are very limited (see Van Tasell, Soli, Kirby, & Widin, 1987). The speech-plus-noise stimuli were created by adding the waveform of each digitized word to that of its matched noise, applying scaling factors to vary the speech-to-noise (*S/N*) ratio while keeping the overall amplitude constant. Six different *S/N* ratios were used: −9.5, −10.7, −12, −13.2, −14.4, and −16.5 dB. All these ratios were well below the identification threshold, according to earlier observations. (A ratio of −7.5 dB was used in the pilot study referred to in the introduction.) Because each word had its own amplitude-matched masking stimulus, any given *S/N* ratio was exactly the same for all words and even for different phonetic segments within each word.

The visual stimuli were presented on a Macintosh computer screen in boldface Geneva font. They subtended an average visual angle of approximately 2.5°. Their presentation was triggered by a tone recorded at the onset of the auditory stimulus, on a second audio channel. The onsets of the spoken words were determined visually on an oscilloscope and were verified auditorily through headphones. The onset was defined as the release of the initial stop consonant in all cases.

Design. There were three experimental groups of 12 subjects each. Each subject was tested at two of the six different *S/N* ratios: relatively high (−9.5 and −10.7 dB), medium (−12 and −13.2 dB), or low (−14.4 and −16.5 dB). At each *S/N* ratio there were 144 trials. Each of the 24 noise and 24 speech-plus-noise stimuli was presented in three different visual conditions: (1) a matching condition (i.e., the same word that was presented auditorily, and/or that was used to generate the noise, was presented in print); (2) a nonmatching condition (i.e., a different word having the same number of phonemes and a similar phonological structure as the word that was

presented auditorily, or was used to generate the noise, was presented in print, e.g., PERSON–BASKET); and (3) a neutral condition in which the visual stimulus was XXXXX.

Procedure and apparatus. The subject was seated in front of the Macintosh computer screen and listened binaurally over Sennheiser headphones. The task consisted of pressing a "yes" key if speech was detected in the noise, and a "no" key if it was not. The dominant hand was used for the "yes" responses. Although the task was introduced as purely auditory, the subjects were requested to attend carefully to the screen as well. They were told in the instructions that, when a word was presented on the screen, it was sometimes similar to the speech or noise presented auditorily, and sometimes not. However they were informed about the equal proportions of "yes" and "no" trials in each of the different visual conditions.

The tape containing the auditory stimuli was placed on a two-channel Crown 800 tape recorder. The verbal stimuli was transmitted to the subject's headphones through one channel, and the trigger tones were transmitted through the other channel to an interface that directly connected to the Macintosh, where they triggered the visual presentation and the computer's clock for reaction time measurements.

The experimental session began with 24 practice trials, after which the first 144 trials were presented in one randomized block, starting with the higher *S/N* ratio. Then there was a 3-min break before the second, more difficult block employing the lower *S/N* ratio.

Results and Discussion

Response percentages. For each subject we determined the percentages of "yes" and "no" responses to speech-plus-noise and noise-only stimuli in each of the three visual conditions. Table 1 shows the average percentages of "yes" responses (i.e., of hits and false alarms); the percentages of "no" responses (misses and correct rejections, respectively) are their complements. There was an extremely high rate of false alarms in all conditions, due to the speech-like envelope of the signal-correlated noise. It is evident that hits decreased and false alarms increased with decreasing *S/N* ratio, as expected. Most interestingly, we see that the percentage of correct detections was higher in the matching print condition than in the other two conditions. This replicates the pilot observations that led to the present experiment. However, the percentage of false alarms was highest in the matching condition also. Apart from the issue of statistical reliability, this raises the question of whether we are dealing here with an increased bias to say "yes" in the matching condition, regardless of whether spectral features were present or not, or whether detectability of spectral properties

TABLE 1
PERCENTAGES OF HITS AND FALSE ALARMS (EXPERIMENT 1)

S/N ratio (dB)	Hits			False alarms		
	Match	No match	XXX	Match	No match	XXX
−9.5	96	92	93	37	19	14
−10.7	97	92	90	41	26	17
−12.0	90	77	77	43	27	21
−13.2	89	80	77	48	29	32
−14.4	74	61	51	56	44	33
−16.5	72	58	45	59	46	34
Average	86	77	72	47	32	25

TABLE 2
DISCRIMINABILITY (*d*) AND BIAS (*b*) INDICES (EXPERIMENT 1)

S/N ratio (dB)	*d*			*b*		
	Match	No match	XXX	Match	No match	XXX
−9.5	2.06	2.36	2.68	1.36	0.27	0.36
−10.7	2.05	2.21	2.32	1.51	0.63	0.22
−12.0	1.42	1.38	1.53	1.06	0.10	−0.04
−13.2	1.46	1.55	1.38	1.24	0.21	0.19
−14.4	0.48	0.44	0.40	0.70	0.05	−0.45
−16.5	0.37	0.30	0.23	0.87	0.09	−0.52
Average	1.30	1.37	1.42	1.13	0.22	−0.04

of speech was in fact increased in the matching condition. To address this question, we turn to an examination of independent discriminability (or sensitivity) and bias indices.

Discriminability and bias indices. Indices of discriminability and bias were computed following the procedures of Luce (1963). Luce's indices were preferred over the standard measures of signal detection theory, d' and β, because they are easier to compute and do not require any assumptions about the shapes of the underlying signal and noise distributions. Moreover, earlier comparisons have shown that results couched in terms of signal detection and Luce indices tend to be very similar (see, e.g., Wood, 1976). The Luce indices, originally named $-\ln\eta$ and $\ln b$, but renamed here for convenience d and b, respectively, are

$$d = \tfrac{1}{2}\ln[p(\text{yes}|S + N)p(\text{no}|N)/ p(\text{yes}|N)p(\text{no}|S + N)]$$

and

$$b = \tfrac{1}{2}\ln[p(\text{yes}|S + N)p(\text{yes}|N)/ p(\text{no}|S + N)p(\text{no}|N)],$$

where $S + N$ and N stand for speech-plus-noise and noise alone, respectively. The discriminability index d assumes values in the same general range as the d' of signal detection theory, with zero representing chance performance. The bias index b assumes positive values for a tendency to say "yes" and negative values for a tendency to say "no".[1]

The average indices are shown in Table 2. Each index was subjected to a three-way analysis of variance with the factors subject group (actually, *S/N* ratio between groups), *S/N* ratio (within groups), and visual condition. The d indices confirm that subjects' performance deteriorated as the *S/N* ratio decreased. At a *S/N* ratio of −16.5 dB, performance was almost at chance level. The main effect of subject group was significant, $F(2,33) = 24.8$, $p = 0.001$, though the main effect of *S/N* ratio within subject groups was not. The latter finding probably represents a practice effect: The more difficult *S/N* ratio always came last in the experimental session and thus received the benefits of practice. The most important result, however, is that subjects' sensitivity was *not* increased in the matching condition. On the contrary, the average d index was lowest in that condition and highest in the neutral condition, though the main effect of visual condition was not significant, $F(2,66) = 1.53$, $p = 0.2$. These differences among visual conditions seemed to be reliable at the two highest *S/N* ratios only, as suggested by a significant interaction of visual condition and subject group, $F(2,66) =$

[1] Given a maximum of 24 responses per subject and condition, values of 0.5 and 23.5 were substituted for response frequencies of 0 and 24, respectively, so as to obtain finite d and b indices.

3.21, $p = 0.02$. We have no explanation for this finding at present. However, our hypothesis that simultaneous matching print might facilitate the detection of the spectral features of speech in noise is clearly disconfirmed.

Turning now to the bias indices, we see a striking difference among the visual conditions: Overall, there was a strong tendency to say "yes" in the matching condition, but little or no bias in the other two conditions. The main effect of visual condition was highly significant, $F(2,66) = 57.1$, $p < 0.001$. In addition, it appears that the overall frequency of "yes" responses decreased with *S/N* ratio in all visual conditions, but this tendency did not reach significance, due to considerable between-subject variability.

The increased frequency of "yes" responses when matching print was present, without a concomitant increase in signal detectability, was obviously caused by the speechlike qualities of the masking noise. In the matching condition, the amplitude envelope of the noise was appropriate for a spoken version of the printed word, and therefore it seemed to the subjects that the word was presented auditorily, whether or not it was in fact hidden in the noise. In retrospect, this explains the subjective impressions of "hearing" words in noise accompanied by print, which led to the present series of experiments. Our data thus reveal that subjects, even when they are not explicitly instructed to do so, automatically detect the correspondence between the amplitude envelope of a nonspeech signal and a sequence of printed letters forming a word, with the consequent illusion of actually hearing the word. This illusion seems akin to the phoneme restoration phenomenon, where surrounding speech context leads subjects to "hear" single phonemes whose acoustic correlates have been replaced by some suitable masking noise (Warren, 1970; Samuel, 1981). The effect revealed in our research suggests that *all* the phonemes in a word may be restored, at least to some extent, when the speech amplitude envelope carried by noise is accompanied by matching print. Since there cannot be a direct connection between the printed letters and the auditory amplitude contour, this might be a top-down effect mediated by a speechlike internal representation of the printed word. Although a global phonetic representation could be envisioned that contains envelope information without a more detailed segmental coding, the striking difference between the matching and the nonmatching visual conditions suggests otherwise: The nonmatching printed words were in fact fairly similar to the matching ones in syllabic stress pattern and phonologic structure, so that a detailed knowledge of the segmental structure would seem to have been necessary to discriminate between their envelopes (see Van Tasell et al., 1987). We conclude, therefore, that printed words are automatically transformed into a detailed phonetic representation, which is probably generated from a more abstract phonological representation stored in the mental lexicon.

Reaction times. Although measures of discriminability and reaction times are usually highly correlated, they may reflect different phases of the cognitive processes involved in the task. While discriminability indices tap into the conscious decision stage, latencies reflect subjects' confidence in reaching their decisions (see Luce, 1986). Therefore, an examination of reaction times may reveal additional information about subjects' processing of the stimuli.

In calculating the average latencies for each subject, outliers beyond two standard deviations from the mean were eliminated. Outliers accounted for less than 2 of the 24 responses per condition, on the average. The average reaction times are presented in Table 3. At the higher *S/N* ratios, there were not enough misses ("no" responses on $S + N$ trials) for meaningful averages to be calculated. Separate analyses of vari-

TABLE 3
REACTION TIMES (EXPERIMENT 1)

S/N ratio (dB)	"Yes" responses					
	Hits			False alarms		
	Match	No match	XXX	Match	No match	XXX
−9.5	667	769	740	920	803	750
−10.7	624	735	724	873	854	775
−12.0	749	843	865	933	1057	1115
−13.2	745	858	834	1015	876	948
−14.4	910	1029	982	1003	1031	1094
−16.5	838	967	951	874	1019	1054
Average	755	867	850	936	940	952
	"No" responses					
	Misses			Correct rejections		
	Match	No match	XXX	Match	No match	XXX
−9.5				914	902	814
−10.7				884	855	808
−12.0		(Insufficient data)		974	1024	896
−13.2				1007	989	876
−14.4	1125	1099	1024	1084	1124	987
−16.5	961	988	904	913	977	909
Average				963	978	882

ance were conducted on hits, false alarms, and correct rejections.

Looking at the hits first, we see that the average latencies increased as the *S/N* ratio decreased across subject groups, $F(2,33) = 6.73$, $p = 0.003$. But no reliable decrease was found within subject groups, probably due to the aforementioned practice effect. As to the effect of visual presentation, we see that the average reaction times were some 100 ms faster in the matching condition than those in the other two conditions. This difference was highly significant, $F(2,66) = 43.08$, $p = 0.001$, and extremely robust: Every single subject showed it, even at the lowest *S/N* ratios.

The false alarm latencies were significantly slower than the hit latencies across all *S/N* ratios, as confirmed in a separate comparison, $F(1,33) = 13.38$, $p = 0.001$. This is consistent with the common finding of slower reaction times for incorrect than for correct responses. However, there was no significant difference among the three visual conditions, nor was there a main effect of *S/N* ratio.

The correct rejection latencies, too, were slower than the hit latencies. They decreased with the *S/N* ratio within subject groups, $F(1,33) = 7.78$, $p = 0.009$, presumably due to practice. The magnitude of that decrease was largest at the lowest *S/N* ratios, which caused a subject group by *S/N* ratio interaction, $F(2,33) = 3.73$, $p = 0.03$. There was no difference between the matching and the nonmatching conditions. However, reaction times were faster in the neutral condition. This effect of visual condition was quite consistent across different *S/N* ratios and was highly significant, $F(2,66) = 23.71$, $p = 0.001$.

The very reliable speeding up of hit responses in the matching visual condition could be explained in terms of the bias to respond "yes" in that condition. However, the false alarms did not show the same decrease even though they were subject to the same response bias (see Table 2). Also, cor-

rect "no" responses might have been expected to show longer latencies in the matching condition. Thus, the reaction time patterns of false alarms and of correct rejections suggest that it was not just the match of print and noise amplitude envelope that caused faster latencies for hits. Rather, it seems that spectral speech information had to be present in order for responses to be speeded up by a match. The faster hit latencies in the matching condition then may reflect, after all, an increase in subjects' sensitivity to the spectral features of the speech signal itself, even though overt detection was not enhanced, and even though the reaction time effect persisted at *S/N* ratios where detectability of the speech approached chance level. Thus, the latencies may tap an earlier level of processing that preceded the conscious decision about presence or absence of the speech signal. This would explain the absence of a similar effect for false alarms, because there was never any signal present for these responses. This interpretation remains speculative, however.

EXPERIMENT 2

The strong response bias caused by the match of print and noise amplitude envelope represents an influence of print on speech perception, a kind of "word restoration" illusion. The main purpose of Experiment 2 was to investigate whether this is a lexical or a prelexical influence. The phonetic representation generated from the print may have been derived from a phonological representation following lexical access or, alternatively, it may have been generated directly from the print via spelling-to-sound conversion rules. One possible method for distinguishing between these two alternatives is to present subjects with nonwords instead of words. Although some authors have argued that nonwords are pronounced by referring to related lexical entries for words (e.g., Glushko, 1979), this route is still less direct than that available for real words. Therefore, if the word restoration effect is lexical in origin, it should be reduced or absent for nonwords. If it is prelexical, on the other hand, it should be obtained for nonwords just as for words. In addition, we wondered whether the intriguing and extremely consistent reaction time facilitation for correct detection of words in the matching condition would be obtained for nonwords as well.

Methods

Subjects. Twelve undergraduate students, all native speakers of English, participated in the experiment for payment.

Stimulus preparation. The stimuli were generated from 24 disyllabic English pseudowords formed by altering one or two letters of real words having the same stress pattern. They had a stop consonant as their initial phoneme, and the number of phonemes ranged from four to six. The written and spoken forms of all nonwords exhibited a regular spelling-to-sound correspondence, according to Venezky (1970); that is, each printed nonword had only one plausible pronunciation—the one spoken. The method for constructing the auditory and the visual stimuli was identical to that of Experiment 1.

Design. Design, procedure, and apparatus of Experiment 2 were identical to those of Experiment 1, except that only one group of subjects was used. Each subject was tested at two *S/N* ratios: −12 and −14.4 dB, in this order.

Results and Discussion

Response percentages. The average percentages of hits and false alarms are presented in Table 4. When these are compared to the results obtained for words with the same *S/N* ratios (Table 1), it is clear that subjects' performance was worse with nonwords: The percentage of hits was lower, and the percentage of false alarms was higher. In addition, the effect of matching print was much smaller in the nonwords: The percentages of correct detections in the matching and nonmatching conditions were

TABLE 4
PERCENTAGES OF HITS AND FALSE ALARMS FOR NONWORDS (EXPERIMENT 2)

S/N ratio (dB)	Hits			False alarms		
	Match	No match	XXX	Match	No match	XXX
−12.0	75	69	64	47	41	33
−14.4	68	69	61	53	47	39
Average	71	69	62	50	44	36

almost identical, and the false alarm percentages showed only a small difference. To examine these effects further we calculated the discriminability and bias indices.

Discriminability and bias indices. The average *d* and *b* indices are presented in Table 5. The *d* indices show that subjects' performance deteriorated as the *S/N* ratio decreased. This main effect was significant, $F(1,11) = 9.3$, $p = 0.01$. At the higher *S/N* ratio, the *d* values were lower than those obtained for words in the previous experiment, suggesting that detection of nonwords was more difficult than that of words. At the lower *S/N* ratio, discriminability was low for both words and nonwords. Apparently, in the present experiment, subjects did not show any effect of practice. As with the words in Experiment 1, the different visual conditions did not affect subjects' sensitivity. The main effect of visual condition was nonsignificant, $F(2,22) = 0.09$.

Analysis of the bias indices revealed a significant effect of visual condition $F(2,11) = 10.0$, $p = 0.001$. Although the direction of the effect was similar to that obtained for words, its size was much smaller for the nonwords. Moreover, a Tukey post hoc analysis revealed that the bias indices in the matching and nonmatching conditions did not differ significantly. In order to assess directly whether the bias effect in the three visual conditions interacted with the lexical status of the stimuli, we conducted a separate analysis in which the nonwords of Experiment 2 and the words of Experiment 1 (for comparable *S/N* ratios) were combined. The interaction of word/nonword and visual condition was significant, $F(2,92) = 6.97$, $p = 0.001$. This outcome demonstrates that the bias effect was indeed different for words and nonwords.

Reaction times. The average reaction times are presented in Table 6. The slow latencies, especially at the higher *S/N* ratio, suggest again that detection of nonwords was more difficult than detection of words. We conducted separate analyses for hits, false alarms, correct rejections, and misses. The pattern of the hits revealed no effect of visual presentation, $F(2,22) = 0.3$, in sharp contrast to the results for words. Thus, for nonwords, matching print did not facilitate correct "yes" responses. Also, neither the effect of *S/N* ratio nor the interaction of *S/N* ratio and visual condition was significant. The significance of the word–nonword difference was again assessed in a separate analysis in which data from Experiments 1

TABLE 5
DISCRIMINABILITY (*d*) AND BIAS (*b*) INDICES (EXPERIMENT 2)

S/N ratio (dB)	*d*			*b*		
	Match	No match	XXX	Match	No match	XXX
−12.0	0.86	0.73	0.76	0.63	0.24	−0.02
−14.4	0.39	0.51	0.56	0.51	0.34	−0.05
Average	0.63	0.62	0.66	0.57	0.29	−0.03

TABLE 6
REACTION TIMES (EXPERIMENT 2)

S/N ratio (dB)	"Yes" responses					
	Hits			False alarms		
	Match	No match	XXX	Match	No match	XXX
−12.0	972	1019	1005	1149	1092	1245
−14.4	1005	999	1017	1020	1071	1081
Average	988	1009	1011	1084	1082	1163
	"No" responses					
	Misses			Correct rejections		
	Match	No match	XXX	Match	No match	XXX
−12.0	1269	1265	1053	1129	1116	1002
−14.4	1120	1070	1055	1048	1091	983
Average	1195	1167	1054	1088	1103	993

and 2 were combined. The interaction of word/nonword and visual condition was indeed significant, $F(2,92) = 3.27$, $p = 0.04$.

The false alarms analysis revealed no significant effect of visual presentation. The analysis of correct rejections, however, did show such an effect, $F(2,11) = 13.8$, $p = 0.001$, due to faster responses in the neutral condition. This unexplained effect is very similar to that found for words. The average reaction times for misses were relatively slow, without any significant effects.

In summary, in contrast to the results previously obtained for words, the bias to say "yes" in the matching condition (the "nonword restoration" effect) was much smaller, and reaction times for correct detections were not faster when the print matched the speech signal. These results support the hypothesis that the word restoration illusion is lexically mediated. Because nonwords are not represented in the mental lexicon, their covert pronunciation is generated either prelexically from the print or indirectly by accessing similar words in the lexicon. Apparently, either process is too slow or too tentative to enable subjects to match the resulting internal phonetic representation to a simultaneous auditory stimulus before that stimulus is fully processed.

One unexpected finding was that overall performance was much worse for nonwords than for words, even though the stimuli were presented at exactly comparable S/N ratios. Had the task required identification of the stimuli, this difference would not have been surprising, since superior recognition performance for real words has been demonstrated in many studies of visual and auditory word perception. However, our subjects could not identify the masked speech, and in most cases they could hardly say if any speech was present at all. How, then, is the poorer performance with nonwords to be explained?

One possibility is that, because words are represented in the lexicon, they are better detected by the perceptual system. This hypothesis was disconfirmed in a recent study reported in detail elsewhere (Repp & Frost, in press): When masked words and nonwords were randomly presented, without simultaneous print, they were detected equally well. Another possibility is that our subjects adopted different perceptual strategies in Experiments 1 and 2. The instantaneous presentation of the print occurred at the beginning of the speech, which unfolded over the next several hundred milliseconds. Almost certainly, processing of the print was completed before that of the

speech. Because of this, and also because only words or nonwords were included in each of the experiments, the subjects always knew in advance whether the auditory stimulus was going to be a word or a nonword. We suspect that this foreknowledge was responsible for the observed difference in performance.

EXPERIMENT 3

The purpose of Experiment 3 was to examine the effect of matching print on detectability of words and nonwords in white noise, instead of signal-correlated noise. In white nose, the amplitude envelope fluctuates little, and randomly, rather than in a speechlike fashion. Therefore, when white noise is used, the "word restoration" effect caused by the match of auditory amplitude envelope and print should disappear completely and, with it, any difference in bias indices.

Even though signal detection theory treats sensitivity and bias as independent parameters, it is conceivable that, in the absence of a strong response bias due to speechlike noise, any effect of matching print on detectability of speech might emerge more clearly, particularly since both spectral and amplitude features can now be utilized by subjects for speech signal detection. Counteracting this possible advantage of using a white noise masker was the possibility that the separation of speech from white noise is much easier and perhaps rests on more peripheral processing of the input. This in turn might reduce any potential top-down effects on subjects' sensitivity. Nevertheless, we wondered whether at least the reaction time difference found for words in Experiment 1 could be replicated in white noise.

Methods

Subjects. Twelve paid undergraduate subjects participated. All were native speakers of English.

Stimuli. The same words and nonwords as those in Experiments 1 and 2 were used. To make the *S/N* ratio comparable for all stimuli, an individual white noise masker was constructed for each speech stimulus as follows: First, white noise produced by a General Radio 1390-A random noise generator was sampled at a rate of 20 kHz and stored in a file. Next, a segment of exactly the same length as the speech was excerpted from that file. Then, 5-ms amplitude ramps were put at the beginning and end of the noise to avoid abrupt onsets and offsets. Subsequently, the average decibel levels of the speech and of the white noise segment were determined, and the white noise (which, as recorded, was from 2 to 7 dB more intense than the speech) was attenuated digitally to exactly the same average amplitude as the speech. The *S/N* ratios, therefore, were specified relative to the average, not the peak, speech signal level. To obtain the speech-plus-noise stimuli, the speech and white noise waveforms were added digitally at a *S/N* ratio of −28 dB, keeping the overall amplitude constant. This ratio was based on previous data (Repp & Frost, in press) and was intended to yield a level of performance around 75% correct.

Design and procedure. Design, procedure, and apparatus were similar to those of Experiments 1 and 2, except that each subject was tested in both the word and the nonword conditions. Half the subjects received the word condition first, and half the nonword condition. The playback level was calibrated for each tape using white noise recorded at the beginning of the tape. The level was set so that the calibration noise registered 0.1 V on a voltmeter, corresponding to 90 dB at the subjects' earphones. The average level of the stimuli was from 2 to 7 dB lower.

Results and Discussion

The average percentages of hits and false alarms are presented in Table 7. The results reveal that matching print did not increase either the hit rate or the false alarm rate for either words or nonwords.

TABLE 7
PERCENTAGES OF HITS AND FALSE ALARMS (EXPERIMENT 3)

	Hits			False alarms		
Stimulus	Match	No match	XXX	Match	No match	XXX
Words	78	75	69	21	21	21
Nonwords	78	78	74	22	22	14

Table 8 presents the *d* and *b* indices. The *d* indices for words and nonwords were very similar, without any significant effects of the different visual conditions. Overall performance in the experiment was relatively good, so the above findings do not result from the inability of subjects to detect the speech in noise.

Table 8 also reveals that, as predicted, the bias effect found in the previous experiments had disappeared. However, there was a significant tendency to give "no" responses in the neutral condition, for both words and nonwords, $F(2,22) = 4.48$, $p = 0.02$. Apparently, the absence of a printed word influenced the subjects' decision criterion.

The reaction times of hits and correct rejections are shown in Table 9. The numbers of false alarms and misses per subject were insufficient for statistical analysis. Hit latencies did not differ significantly across the different visual conditions, or between words and nonwords. This result is consistent with all of the above findings. Reaction times for correct rejections revealed a significant interaction of visual condition and stimulus type, $F(2,22) = 10.4$, $p = 0.001$. However, this interaction is uninterpretable because the "match" on noise-only trials concerned solely the average amplitude and duration of the noise. It seems unlikely that such "matches" had any influence on subjects' responses. The variability in reaction times may just reflect differential responses to the "matching" and "nonmatching" visual word stimuli.

In summary, as we hypothesized, when the masking noise did not include the envelope information, the effect of matching print on response bias disappeared. However, the detection of speech was not enhanced by matching print. Moreover, matching print did not affect reaction times of correct word detections, perhaps because the detection of speech in white noise rests on different criteria than detection in signal-correlated noise.[2] Speech detection in white noise may be a superficial task that does not require a detailed analysis of the auditory input, and therefore is "out of reach" for top-down effects. The absence of an overall performance difference between words and nonwords is also consistent with this interpretation.

GENERAL DISCUSSION

In the present study we used a signal detection task to investigate whether simultaneous presentation of matching print can affect the detectability of speech in noise. In Experiment 1 we found no evidence for an enhancement of subjects' sensitivity to the spectral features of the speech. However, the reaction time analysis revealed that subjects' confidence in their decisions was increased by the match, but only when spectral information was indeed present in the noise. There was also a strong bias toward "yes" responses, caused by the correspondence of print and noise amplitude envelope. From Experiment 2 we learned that printed nonwords elicit a much smaller bias effect and show no facilitation of reac-

[2] Note also that the *S/N* ratio for white noise was much lower than that for signal-correlated noise at a similar performance level. The ratios are not exactly comparable, however, because they have a different reference. *S/N* ratios would have been more similar if the white noise had been specified with reference to peak, rather than average, speech signal levels. (See Horii, House, and Hughes, 1971.)

TABLE 8
DISCRIMINABILITY (*d*) AND BIAS (*b*) INDICES (EXPERIMENT 3)

	d			*b*		
Stimulus	Match	No match	XXX	Match	No match	XXX
Words	1.36	1.30	1.25	0.02	−0.12	−0.38
Nonwords	1.34	1.32	1.63	−0.01	0.05	−0.45

tion times. Finally, Experiment 3 demonstrated that there is no influence of print even for words when white noise is used as the masking sound.

The results of Experiment 1, together with earlier phenomenological observations, suggest that, when signal-correlated noise is employed, the presentation of matching print generates a perceptual illusion: Subjects believe they hear speech, even when no speech is present in the noise. That the bias is mediated by the amplitude envelope of the noise was confirmed in Experiment 3, where the effect was totally absent with white noise.

The bias effect suggests that the printed words were immediately recoded into an internal phonetic form. In order for subjects' responses to be affected by the match between print and noise amplitude envelope, the information generated from the visual and the auditory modalities must have been in the same internal metric. The amplitude envelope of the auditory stimulus was almost certainly insufficient to generate a detailed abstract phonologic code that could have been compared to phonologic information accessed from print. Therefore, it is the print that must have been converted internally into a phonetic representation. What our findings teach us is that a phonetic code is generated from printed words *automatically*, even when the task does not require it. After all, our subjects were never instructed to match the print to the auditory stimuli and were specifically informed of the equal distribution of speech-plus-noise and noise-only trials in the different visual conditions. It is unlikely that amplitude envelopes are stored as such in the lexicon, since they are contingent on phonetic structure. Their availability from print implies that a segmental phonetic representation is generated. Thus, our results provide further confirmation for the notion of obligatory and fast phonetic coding in reading, suggested by a large literature on visual word perception but sometimes challenged by those who find evidence for rapid lexical access based on orthography alone. We suspect that phonetic coding takes place regardless of what information gets to the lexicon first.

How was it possible for matching print to influence reaction times only for correct detections, without an apparent increase in sensitivity to spectral speech features? A possible explanation of this pattern of results is that matching print had a confirmatory effect at a processing stage *following* the extraction of partial spectral features from the noise. Thus, the subjects' confidence in correct detections was increased in the matching condition without any ac-

TABLE 9
REACTION TIMES (EXPERIMENT 3)

	"Yes" responses			"No" responses		
	Hits			Correct rejections		
Stimulus	Match	No match	XXX	Match	No match	XXX
Words	706	731	717	818	891	804
Nonwords	720	722	762	909	882	857

tual increase in the amount of spectral information extracted.

The strong bias effect and the facilitation of reaction time were not obtained for nonwords. Therefore the influence of printed words on speech processing appears to be lexically mediated. That is, the internal phonetic representation is probably generated from a more abstract phonological code stored in the lexicon, rather than by applying spelling-to-sound conversion rules. Nonwords are clearly at a disadvantage in such process.[3] Whether this disadvantage consists merely of a longer processing time could be tested by presenting the visual information somewhat in advance of the onset of the auditory stimulus, so that there is sufficient time to recast nonwords into an internal phonetic code. The observed difference between words and nonwords might then disappear. At this time, we can only conclude that the covert naming of printed nonwords is not as efficient as that of words, which certainly agrees with previous results obtained in overt naming tasks.

No bias or facilitation of reaction time was obtained for words or nonwords in white noise. Whereas the signal-correlated noise masker forced subjects to extract speech-specific spectral information, essentially phonetic features, the white noise masker permitted use of simple auditory strategies: Any deviation from the random noise background, whether speechlike or not, could be used in the decision process. Given the very low *S/N* ratios used, the information on which the subjects' decisions were based probably was not speechlike enough to interact with phonetic top-down information.

[3] The claim that the generation of phonetic structure from print is faster postlexically than prelexically does not imply that fast lexical access for words is achieved by a "visual route." Access to the lexicon may be achieved when sufficient phonologic information for determining a specific lexical entry has accumulated prelexically. However, such information may not be sufficient for the generation of a detailed phonetic structure.

In conclusion, we find support in our results for an interactive view of the processes of visual and auditory word perception, even though the early auditory processes of spectral feature extraction appear to be impermeable to top-down influences. The interaction of the visual and auditory word processing systems (in the direction we investigated, viz., from visual to auditory) seems to take place because of a rapid recoding of printed words into internal speech comparable in all respects to the output of the auditory phonetic module.

APPENDIX: WORDS AND NONWORDS USED IN THE EXPERIMENTS

PUBLIC	TEAMON
BODY	DEEMY
CLOSET	KETTER
BABY	BAXI
DOLLAR	DALIK
PICTURE	PIRTON
PAPER	PAMET
TEMPLE	TRISIN
PERSON	TILBER
PARENT	BEALTY
CARGO	PINOW
TABLE	TARNET
CANYON	TONKOR
CORNER	DORIT
TOTAL	PROSOR
CANVAS	BOONTER
DANGER	QUEMPLE
KITCHEN	BOTCHEN
PUPIL	PUNIL
DIMPLE	TUNY
PANIC	PAGER
PRISON	PROSOR
GARDEN	GASNET
PENCIL	CALVAS

REFERENCES

COLE, R. A., & JAKIMIK, J. (1978). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–164). Hillsdale, NJ: Erlbaum.

FRAUENFELDER, U. H., & TYLER, L. K. (1987). The process of word recognition: An introduction. *Cognition*, 25, 1–20.

FROST, R., & KATZ, L. (Submitted). Orthographic depth and the interaction of visual and auditory processing in word recognition.

GLUSHKO, R. J. (1979). The organization of activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 674–691.

HORII, Y., HOUSE, A. S., & HUGHES, G. W. (1971). A masking noise with speech-envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America,* **49,** 1849–1856.

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

KUCERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology.* New York: Wiley.

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

MARSLEN-WILSON, W. D., & WELSH, A. (1978). Processing interaction during word recognition in continuous speech. *Cognitive Psychology,* **10,** 29–63.

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

MCCUSKER, L. X., HILLINGER, M. L., & BIAS, R. G. (1981). Phonologic recoding and reading. *Psychological Bulletin,* **89,** 217–245.

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature,* **264,** 746–748.

REPP, B. H., & FROST, R. (in press). Detection of words and nonwords in two kinds of noise. *Journal of the Acoustical Society of America.*

SAMUEL, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General,* **110,** 474–494.

SCHROEDER, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America,* **43,** 1735–1736.

SUMMERFIELD, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 3–52). Hillsdale, NJ: Erlbaum.

VAN TASELL, D. J., SOLI, S. D., KIRBY, V. M., & WIDIN, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America,* **82,** 1152–1161.

VENEZKY, R. L. (1970). *The structure of English orthography.* The Hague: Mouton.

WARREN, R. M. (1970). Perceptual restoration of missing speech sounds. *Science,* **167,** 392–393.

WOOD, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America,* **60,** 1381–1389.

(Received February 9, 1988)

(Revision received June 1, 1988)