781

Cognition, 39 (1991) 195–214

Phonetic recoding of print and its effect on the detection of concurrent speech in amplitude-modulated noise*

RAM FROST

Hebrew University, Jerusalem

Received October 23, 1990, final revision accepted April 29, 1991

Abstract

Frost, R., 1991. Phonetic recoding of print and its effect on the detection of concurrent speech in amplitude-modulated noise. Cognition, 39: 195–214.

When an amplitude-modulated noise generated from a spoken word is presented simultaneously with the word's printed version, the noise sounds more speechlike. This auditory illusion obtained by Frost, Repp, and Katz (1988) suggests that subjects detect correspondences between speech amplitude envelopes and printed stimuli. The present study investigated whether the speech envelope is assembled from the printed word or whether it is lexically addressed. In two experiments subjects were presented with speech-plus-noise and with noise-only trials, and were required to detect the speech in the noise. The auditory stimuli were accompanied with matching or non-matching Hebrew print, which was unvoweled in Experiment 1 and voweled in Experiment 2. The stimuli of both experiments consisted of high-frequency words, low-frequency words, and non-words. The results demonstrated that matching print caused a strong bias to detect speech in the noise when the stimuli were either high- or low-frequency words, whereas no bias was found for non-words. The bias effect for words or non-words was not affected by spelling to sound regularity; that is, similar effects were obtained in the voweled and the unvoweled conditions. These results suggest that the amplitude envelope of the word is not assembled from the print. Rather, it is addressed directly from the printed word and retrieved from the mental lexicon. Since amplitude envelopes are contingent on detailed phonetic structures, this outcome suggests that representations of words in the mental lexicon are not only phonological but also phonetic in character.

---

*This work was supported in part by a grant awarded to the author by the Basic Research Foundation administered by the Israel Academy of Science and Humanities, and in part by the National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories. The author is indebted to Liora Yancovitch for her help in conducting the experiments, and to Benny Shanon, Bruno Repp, and three anonymous reviewers for their comments on earlier drafts of this paper. Correspondence concerning this article should be addressed to Ram Frost, Department of Psychology, The Hebrew University, Jerusalem, 91905 Israel.

Introduction

It is generally assumed that the processing of words in the visual and auditory modalities differs in the initial phase because of different input characteristics, but converges at later stages. Hence, findings regarding the influence of orthographic information on the perception of speech, and findings showing how spoken information affects visual word perception, may suggest how print and speech are integrated in the mental lexicon. The present study is concerned with a special form of interaction between the visual and auditory modalities during word recognition. It discusses the possible origins of an illusion of hearing speech in noise caused by simultaneous presentation of printed information.

The convergence of printed and spoken stimuli representations during processing has been previously demonstrated in unimodal studies. It has been shown that lexical decisions to spoken words are facilitated if successive words share the same spelling (Jakimik, Cole, & Rudnicky, 1980). Similarly, Hillinger (1980) has shown that priming effects with printed words were enhanced when primes and targets were phonemically similar. However, the influence of one modality on processing in the other modality can be shown more directly in cross-modal studies. It has been established that printed words can prime lexical decisions to spoken words and vice versa (Kirsner, Milech, & Standen, 1983; Hanson, 1980). Similarly, using the naming task, Tanenhaus, Flanigan, & Seidenberg (1980) have demonstrated a visual–auditory interference in a Stroop paradigm. These results were interpreted to show that reading and listening share one lexicon, which allows identical messages to be understood in the two modalities in the same way.

Stronger but more controversial evidence concerning the interaction of the visual and the auditory modalities comes from studies demonstrating cross-modal influence occurring before the completion of input analysis. According to a strongly interactive view, some or all stages of the *perceptual* process in one modality may be influenced by activation in the other modality. For example, it has been suggested that automatic grapheme-to-phoneme activation might occur prior to word recognition, thereby affecting the process of auditory lexical access through sub-lexical activation in the visual modality (e.g., Dijkstra, Schreuder, & Frauenfelder, 1989; Frost & Katz, 1989). Dijkstra et al. (1989) have shown that a visual letter prime can facilitate the auditory detection of a vowel in a syllable. Similarly, Layer, Pastore, and Rettberg (1990) have reported results showing faster identification of an initial auditory phoneme when congruent visual information was presented simultaneously.

Perceptual cross-modal influences can be shown at levels higher than graphemes and phonemes. In a recent study, Frost, Repp, and Katz (1988) have reported an auditory illusion occurring when printed words and masked spoken words appear simultaneously. Subjects were presented with speech-plus-noise and with noise-only trials, and were required to detect the masked speech in a signal detection paradigm. The auditory stimuli accompanied by print which either matched or did not match the masked speech. Since the noise used in this experiment was amplitude-modulated, (i.e., the spoken word was masked by noise with the same amplitude envelope), when a printed word matched the spoken word, it also matched the amplitude envelope of the noise generated from it. Frost et al. (1988) have shown that, whether speech was indeed present in the noise or not, subjects had the illusion of hearing it in the noise when the printed stimuli matched the auditory input. These results demonstrate that subjects automatically detected a correspondence between noise amplitude envelopes and printed stimuli when they matched. The detection of this correspondence made the amplitude-modulated noise sound more speechlike, causing a strong response bias. This effect was extremely reliable and appeared for every subject tested. The bias effect did not appear when the printed words and the spoken words from which the amplitude envelopes were generated were merely similar in their syllabic stress pattern, or phonological structure. These results suggest that the printed words were recoded into a very detailed, speechlike, phonetic representation that matched the auditory information, thereby causing the illusion.

One important finding reported by Frost et al. (1988) relates to the processing of non-words. When the printed and spoken stimuli were pseudo-words (non-words which were phonotactically regular), the bias to hear speech in the noise in the matching condition was much smaller. This result is of special interest because subjects could not identify the masked spoken stimuli, and therefore were unaware that they consisted of non-words. Nevertheless, they could not detect a correspondence between a printed letter string and its amplitude envelope if it was not a legal word. One possible interpretation of this outcome is that in contrast to words, the covert pronunciation of non-words is generated either pre-lexically 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.

However, a more radical interpretation of the words–non-words differences can be suggested. It is possible that amplitude envelopes are stored as holistic patterns in the lexicon, and are addressed automatically by printed words. According to this interpretation, the bias effect could not have been obtained for non-words, because non-words are not represented in the mental lexicon, and their printed forms could not have addressed any stored amplitude envelope. It is important to explore this hypothesis further since it has direct relevance to models concerned with the representations of spoken words in the mental lexicon, and with models of visual lexical access. Models of spoken word recognition often assume that representations of words in the lexicon are phonological in nature, and that the contact representations generated from the speech wave are abstract linguistic units like phonemes and syllables (see Frauenfelder & Tyler, 1987, for a review). According to the above interpretation, however, representations of spoken words are maximally rich, consisting not only of abstract linguistic units, but also of detailed phonetic information such as spectral templates, and amplitude envelopes. Amplitude envelopes cannot be considered phonological representations because they do not provide the explicit phonemic or syllabic structure of the word. Rather, they retain some speechlike features and convey mostly prosodic and stress information. A similar non-phonological approach to the mental lexicon was advocated by Klatt in his LAFS (lexical access from spectra) model (Klatt, 1979; see Klatt, 1989, for a review; see also Gordon, 1988; Jusczyk, 1985).

This issue is also relevant to current discussions concerning the processing of printed words. Models of visual word perception are in disagreement concerning the extent of phonological recoding during printed word recognition (e.g., Seidenberg, 1985; Van Orden, 1987). One class of models assumes that phonological codes are generated automatically following visual presentation and mediate lexical access (Perfetti, Bell, & Delaney, 1988; Van Orden, Johnston, & Hale, 1988; and see Van Orden, Pennington, & Stone, 1990, for a review). In contrast, it has been suggested that phonological codes are seldom generated during visual word recognition, and that with the exception of very infrequent words, printed words activate orthographic units that are directly related to meaning in semantic memory (e.g., Seidenberg, 1985; Seidenberg, Waters, Barnes, & Tanenhaus, 1984). Thus, results demonstrating that a visual presentation of a printed word produces a detailed phonetic representation that includes the word's amplitude envelope, even when the experimental task does not require it, provide support for automatic and rapid phonetic recoding in silent reading.

The aim of the present study was to examine further the hypothesis that amplitude envelopes representations of spoken words are not assembled pre-lexically from the print, but are stored holistically in the mental lexicon, and are addressed directly and automatically by matching printed words following lexical access. The generation of a phonetic representation from the print can theoretically be achieved through a pre-lexical process that maps representation of graphemes into representation of phonemes by applying grapheme–phoneme correspondence rules, and subsequently by transforming the abstract phonological structure into a detailed phonetic representation for silent or overt reading. This process has been often suggested to characterize the naming of novel words or of non-words (e.g., Coltheart, 1978). Note that whether the phonological and phonetic structures are derived by applying grapheme–phoneme correspondence rules (Venezky, 1970), or by analogy (Glushko, 1979) is irrelevant in the present context, since both procedures assume that the phonological code is generated *prior* to the selection of a lexical candidate (i.e., prior to lexical access). In contrast to this account, the hypothesis forwarded in the present study suggests that possible differences in bias between words and non-words do not result from the relative speed or ease with which the graphemic structure is transformed pre-lexically into a phonetic code. Rather, they emerge because printed words address a maximally rich lexical representation which contains, among other things, the amplitude envelope of the spoken word. Non-words, on the other hand, are not represented in the mental lexicon, and therefore cannot address their amplitude envelope.

For this purpose, the present study employed the speech detection task proposed by Frost et al. (1988) and examined whether the bias effect caused by matching print depends on the speed of print processing and on spelling-to-sound regularity, or whether it has a lexical origin. Spelling-to-sound regularity and the speed of generating phonological codes were manipulated by using word frequency and the unique characteristics of the Hebrew orthography.

In the two experiments reported here, subjects were presented auditorily with speech-plus-noise or with noise-only trials, simultaneous with a visual presentation of printed Hebrew words. In Hebrew, letters represent mostly consonants, while vowels can optionally be superimposed on the consonants as diacritical marks. Like other Semitic languages, Hebrew is based on word families derived from tri-consonant roots. Therefore, many words share a similar or an identical letter configuration. If the vowel marks are absent, a single printed consonantal string usually represents several different spoken words. Thus, in its unvoweled form, the Hebrew orthography is considered a very deep orthography: it does not convey to the reader the full phonemic structure of the printed word, and the reader is often faced with phonological ambiguity.[1] In contrast, the voweled form is a very shallow writing system.

---

[1] A demonstration of this form of ambiguity may be portrayed in English by the following example: the consonantal string "bttr" may be read as "better", "butter", "bitter" or "batter", which are meaningful words. In addition, many other vowel configurations could be added to the consonants to form non-words. The Hebrew reader is faced with this form of phonological ambiguity regularly in the unvoweled orthography. The addition of the diacritical marks specifies uniquely one phonological alternative.

The vowel marks convey the missing phonemic information, making the printed word phonemically unequivocal (but see Frost, in press, for a discussion).

Several studies in Hebrew have established that the presentation of unvoweled print encourages the use of orthographic codes to access the lexicon. In order to assign a correct vowel configuration to the printed consonants to form a valid word, readers of Hebrew have to draw upon their lexical knowledge. The complete phonological structure of the printed word can only be retrieved post-lexically, after one word candidate has been accessed. (Bentin, Bargai, & Katz, 1984; Frost, Katz, & Bentin, 1987). In contrast, the explicit presentation of vowel marks provides the reader with the complete phonemic structure of the word (or non-word). Because the voweled orthography is characterized by grapheme-to-phoneme regularity, the diacritical marks enable the generation of a pre-lexical phonological code by using simple spelling-to-sound conversion rules. This special characteristic of the Hebrew orthography was exploited in order to investigate whether the bias effect caused by matching print on speech detection in noise is affected by the print's phonological transparency. Specifically, we examined whether the bias effect is dependent on the presentation or the omission of vowel marks.

Experiment 1

In Experiment 1 subjects were presented with high- and low-frequency spoken words, as well as with non-words, which were masked by noise with the same amplitude envelope. In addition, the noises were presented alone. The subjects' task consisted of deciding in each trial whether speech was present in the noise, or whether there was noise only. Simultaneous with the auditory presentation, a printed unvoweled Hebrew letter string appeared on a computer screen. Sometimes the printed word or non-word matched the auditory stimulus, and sometimes it did not. In each of these experimental conditions it was determined whether the print caused a bias to hear speech in the noise.

The purpose of this experiment was three-fold: First, to examine whether the bias effect obtained in the shallower English orthography can be obtained in the deeper Hebrew orthography. If the effect depends on the speed of generating amplitude envelopes pre-lexically from the print by using spelling-to-sound conversion rules, then the unvoweled Hebrew is at a clear disadvantage. It does not convey explicitly to the reader the full phonemic information necessary for the construction of the amplitude envelope. Although the vowel information can be retrieved from the lexicon following visual lexical access, this process is slower to develop. Indeed, a multilingual comparison of naming latencies (Frost et al., 1987) revealed that naming in unvoweled Hebrew is slower than naming latencies in shallower orthographies like English and Serbo-Croatian. Moreover, in contrast to English and Serbo-Croatian, naming latencies in Hebrew were found to be slower than lexical decisions. This is because the phonemic structure necessary for naming is not conveyed directly by the print, but retrieved from the lexicon (Frost et al., 1987).

Another factor which affects the speed of generating phonetic codes from print is word frequency. Hence, the second aim of Experiment 1 was to examine whether the bias effect, if obtained, depends on word frequency, or merely on word lexicality. If our previous differences in bias between words and non-words resulted from the speed by which the printed words and non-words were transformed into a phonetic structure, then one should expect a stronger effect of bias for high-frequency words relative to low-frequency words. This is because it is easier to retrieve the phonetic structure of high-frequency words (as reflected by faster reaction times (RTs) for these words in the naming task). If, on the other hand, the origin of the bias effect is purely lexical, then a bias should be obtained for all words, whether frequent or non-frequent, but not for non-words.

Finally, the third aim of the experiment was to examine the bias effect in a mixed design of words and non-words. Note that in the original study reported by Frost et al. (1988) we employed a blocked design. One serious handicap with our previous blocked design was that subjects knew in advance whether the auditory stimuli were words or non-words. This might have encouraged the adoption of different strategies for words and for non-words, thereby causing the differences we obtained in the bias effect. In a mixed design such uniform strategy cannot be adopted. Thus, if our previous results were caused by this methodological factor, then no significant differences in bias between words and non-words should emerge in the present mixed design, and non-words would show the effect as well.

Methods

Subjects

Twenty-four undergraduate students, all native speakers of Hebrew, participated in the experiment for course credit or for payment.

Stimulus preparation

The stimuli were generated from 24 disyllabic words and 12 disyllabic non-words that had a stop consonant as their initial phoneme. The number of phonemes for all stimuli was either four or five. The 24 words consisted of 12 high-frequency words and 12 low-frequency words. Because there are no reliable sources of standard objective word frequency counts in Hebrew, subjective frequencies were assessed by averaging the ratings of 50 subjects on a 1 (least frequent) to 7 (most frequent) scale. The mean ratings of the high- and the low-frequency words were 5.3 and 3.1, respectively. The 24 words were all unambiguous in unvoweled print; that is, their orthographic form represented only one lexical entry. Thus, each letter string could be read as a meaningful word in only one way, by assigning to the consonant one specific vowel configuration. The non-words were, in fact, pseudo-words, that were constructed by altering one or two phonemes of real words. All non-words conformed to the phonotactic rules of the Hebrew language.

The auditory stimuli were originally spoken by a male native speaker in an acoustically shielded booth and recorded on an Otari MX5050 tape-recorder. The speech was digitized at a 20-kHz sampling rate. From each digitized word, a noise stimulus with the same amplitude envelope was created by randomly reversing the polarity of individual samples with a probability of 0.5 (Schroeder, 1968). This 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 the matched noise, adjusting their relative intensity to yield a signal-to-noise ratio of −10.7 dB.

Each digitized stimulus was edited using a waveform editor. The stimulus onset was determined visually on an oscilloscope and was verified auditorily through headphones. A mark tone was then inserted at the onset of each stimulus, on a second track that was inaudible to the subjects. The digitized edited stimuli were recorded at three-second intervals on a two-track audiotape, one track containing the spoken words while the other track contained the mark tones. The purpose of the mark tone was to trigger the presentation of the printed stimuli on a Macintosh computer screen.

Design

Each of the high-frequency words, low-frequency words and non-words was presented in two auditory forms: (1) speech-plus-noise trials, in which the spoken stimulus was presented masked by noise; (2) noise-only trials, in which the noise was presented by itself without the speech. Each of these auditory presentations was accompanied by two possible visual presentations: (1) a matching condition (i.e., the same word or non-word that was presented auditorily and/or that was used to generate the amplitude-modulated noise, was presented in print); (2) a non-matching condition (i.e., a different word or non-word, having the same number of phonemes and a similar phonological structure as the word or non-word presented auditorily, or that was used to generate the noise, was presented in print). Thus, there were four combinations of visual/auditory presentations for each word or non-word, making a total of 144 trials in the experiment.

Procedure and apparatus

Subjects were seated in front of a Macintosh SE computer screen (9-inch diagonal, screen size), and listened binaurally over Sennheiser headphones at a comfortable intensity. The subjects sat approximately 70 cm from the screen, so that the stimuli subtended a horizontal visual angle of 4° on the average. A bold Hebrew font, size 24, was used. 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 always 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 or a non-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 Otari MX5050 tape-recorder. The verbal stimuli were 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.

The experimental session began with 24 practice trials, after which the 144 experimental trials were presented in one block.

Results and discussion

The indices of bias in the different experimental conditions were computed following the procedure suggested by Luce (1963). Results computed according to Luce's procedure tend to be very similar to results produced by the standard signal detection computations (e.g., Wood, 1976). However, Luce's indices do not require any assumptions about the shapes of the underlying signal and noise distributions, and are easier to compute relative to the standard measures of signal detection theory. The Luce indices of bias and sensitivity, originally named $\ln b$ and $\ln \eta$, but renamed here for convenience b and d are:

$$b = (1/2)\ln[p(\text{yes}/\text{s}+\text{n})p(\text{yes}/\text{n})/p(\text{no}/\text{s}+\text{n})p(\text{no}/\text{n})]$$

and

$$d = (1/2)\ln[p(\text{yes}/\text{s}+\text{n})p(\text{no}/\text{n})/p(\text{yes}/\text{n})p(\text{no}/\text{s}+\text{n})]$$

where s + n and n stand for speech-plus-noise and noise only, respectively. The index *b* assumes positive values for a tendency to say "yes" and negative values for a tendency to say "no". For example, according to the above formula, in order to obtain an average *b* of +0.5, the subject must generate on the average 60% more positive responses than negative ones. The index *d* assumes values in the same general range as the d' of signal detection theory, with zero representing chance performance.

The average values for the bias indices in each experimental condition are shown in Table 1 (top). There was a bias to say "yes" in the matching condition for high-frequency and for low-frequency words. whereas there was no bias in the non-matching condition. The bias effect found for high-frequency words was not stronger than that for low-frequency words. In fact the opposite pattern was obtained. In contrast to the high- and the low-frequency words, there was no bias to say "yes" for non-words in the matching condition. There was, however, a bias to say "no" in the non-matching condition.

Table 1. *Bias indices* (b), *and (standard error of the means) for high-frequency words, low-frequency words and non-words, when matching and non-matching print is presented simultaneously with masked speech. Print is presented* unvoweled. *The top* b *indices were averaged for all subjects, whereas the bottom* b *indices were averaged for the 12 subjects with the highest detectability scores* (d)

	High-frequency words	Low-frequency words	Non-words
Match	0.55 (0.14)	0.94 (0.18)	−0.19 (0.14)
No match	−0.11 (0.12)	0.02 (0.14)	−0.48 (0.13)
Average *d* = 0.15 (*n* = 24)			
Match	0.57 (0.20)	0.99 (0.19)	−0.20 (0.15)
No match	−0.35 (0.17)	−0.15 (0.13)	−0.67 (0.17)
Average *d* = 0.32 (*n* = 12)			

The bias indices were subjected to a two-way analysis of variance with the factors of word type (high-frequency words, low-frequency words and non-words) and visual condition (matching print, non-matching print). The main effects of word type and visual condition were significant ($F(2,46) = 17.3$, MSe = 0.48, $p < .001$, and $F(1,23) = 22.0$, MSe = 0.64, $p < .001$, respectively). The two-way interaction was also significant ($F(2,46) = 6.5$, MSe = 0.19, $p < .003$). A Tukey post-hoc analysis revealed that the differences in bias between either type of words and between the non-words were reliable, as well as the difference between the high- and the low-frequency words ($p < .05$). The apparent greater bias to say "no" in the non-matching condition relatively to the non-matching condition for the non-words was not significant.

The average d in the experiment was 0.15. Hence, the signal-to-noise ratio which was employed in the experiment resulted in a very low level of detection.[2] In order to ensure that the obtained pattern of bias was not affected by the low detection level, the subject sample was split in half, and the average b indices were recomputed for those subjects with highest d. The average b for this sample in the different experimental conditions are presented in Table 1 (bottom), and confirm that the bias was unaffected by the level of detection. This outcome is in accordance with results presented by Frost and his colleagues showing significant bias effects over a wide range of signal-to-noise ratios.[3]

The data of Experiment 1 thus reveal that the bias in the visual matching condition was obtained even in the unvoweled Hebrew orthography. However, similar to our previous study, this effect can be demonstrated only when legal words are presented in the visual modality. The bias effect was not reduced when the printed words were relatively infrequent. This suggests that the speed by which the phonetic structure of the word is retrieved does not

---

[2] The discriminability indices obtained in the present experiments were lower than those obtained by Frost et al. (1988), with a comparable signal-to-noise ratio. This difference may possibly be attributed to differences in the spoken stimuli from which the envelopes were generated. In the present study the spoken stimuli were recorded by a male speaker, whereas Frost et al. (1988) employed stimuli recorded by a female speaker. The detection of speech in amplitude-modulated noise is achieved by perceiving local spectral peaks that rise above the flat spectral level represented by the masking noise. Such peaks are more salient with a female speaker because of the higher frequencies that are characteristic to female voices.

[3] Although Frost et al. (1988) showed significant bias effects over a wide range of signal-to-noise ratios, they found reduced bias indices at the lowest ratios. Hence, it is possible that the bias values obtained in the present study were lower than those obtained by Frost et al. (1988), because of the lower level of detection in the present experiment. Note, however, that the interpretation of the results is unaffected by the overall bias values, since it is concerned with the *differences* in bias between the matching and the non-matching conditions.

affect the illusion of hearing speech in the noise in the matching condition.[4] The unexpected stronger effect of bias obtained for the low-frequency words may possibly be related to the phonetic features of the words employed. This possibility will be further considered in the General Discussion.

The most significant outcome of the experiment is that there was no bias in the matching condition for non-words. Since in the present study a mixed design was employed, this effect cannot be attributed to a uniform "set" strategy adopted for the non-words. Although there was a greater tendency to say "no" when non-words appeared in the non-matching condition relative to the matching condition, this tendency was not found to be statistically reliable. These results suggest, then, that in contrast to words the presentation of printed non-words did not easily invoke a phonetic representation which could be compared to the amplitude envelopes presented auditorily, Hence, the outcome of Experiment 1 lends support to the hypothesis that the bias to say "yes" in the matching condition for words only, regardless of their frequency, results from the automatic retrieval of their amplitude envelopes from the lexicon. This process does not appear to be affected by factors related to the speed of generating a phonetic code.

Experiment 2

One possible criticism of the results of Experiment 1 is that in the unvoweled Hebrew orthography the phonemic structure of printed words can be retrieved from the mental lexicon following visual access. In contrast, the phonemic structure of non-words cannot be determined unequivocally, since the printed consonants do not specify how exactly a non-word should be read. It might be argued that this caused the different pattern of bias found for words relatively to non-words. According to this interpretation, the amplitude envelopes of words were not stored as such in the lexicon, but generated on-line from more abstract phonological or phonetic structures

---

[4] Throughout this paper the assumption is made that the processing of printed low-frequency words is slower than the processing of printed high-frequency words. This assumption is supported by the well-documented frequency effect in visual word recognition, but was not directly examined in the present study (subjects were not required to convey their decisions with any time constraints). Decision latencies in the speech detection task do not reflect exclusively the speed of processing the printed words, but also the complexity of processing the auditory stimuli. Hence, the monitoring of reaction times in this task does not necessarily portray the speed of processing the printed information. Note, however, that the set of stimuli employed in Experiments 1 and 2 is a subset of the stimuli examined by Frost, Katz, and Bentin (1987) in the lexical decision and the naming tasks. The frequency effect obtained by Frost et al. (1987) was over 100 ms, supporting the assumption that the processing of the low-frequency words was indeed slower.

which were retrieved *post-lexically* for the words. Because non-words are not represented in the lexicon, and because the complete phonetic structure of the non-words was not specified by the unvoweled print, no bias was obtained for non-words.

In order to ascertain that this factor did not affect our previous findings, in Experiment 2 the effect of bias was measured when the printed stimuli were voweled. By adding the diacritical vowels marks to the consonants, the Hebrew orthography is as shallow as other orthographies which have a clear and unequivocal mapping of spelling-to-sound (e.g., Serbo-Croatian). The marks convey the full phonemic information that is necessary to produce a pre-lexical phonological code for both words and non-words. Therefore, the explicit presentation of vowels eliminates the superiority of words over non-words in regard to phonological and phonetic processing: Phonological recoding of both words and non-words can easily and unequivocally occur through a fast pre-lexical process by applying grapheme-to-phoneme correspondence rules, and a phonetic representation that includes the word's amplitude nvelope may be generated subsequently from the pre-lexical phonological representation. If, indeed, an amplitude envelope can be formed on-line such pre-lexical representations, then the addition of vowel marks should produce a bias effect for non-words as well as for words in the matching condition.

Method

Subjects

Twenty-four undergraduate students, all native speakers of Hebrew, participated in the experiment for course credit or for payment. None of the subjects participated in Experiment 1.

The design, procedure, and apparatus were identical to Experiment 1, except that the printed words and non-words were voweled by adding their diacritical marks.

Results and discussion

The bias indices are presented in Table 2. As in Experiment 1, there was a bias to say "yes" in the matching condition for high- and for low-frequency words, but no positive bias whatsoever for non-words. There was no positive bias in the non-matching condition for words. However, similar to the pattern obtained in Experiment 1, there was a bias to say "no" in the non-matching condition for non-words. The b indices were subjected to a two-way ANOVA with the factors of word type and visual presentation. The main effects of

Table 2. *Bias indices (b), and (standard error of the means) for high-frequency words, low-frequency words and non-words, when matching and non-matching print is presented simultaneously with masked speech. Print is presented* voweled

	High-frequency words	Low-frequency words	Non-words
Match	0.67 (0.16)	0.84 (0.18)	−0.00 (0.15)
No match	−0.16 (0.15)	−0.06 (0.15)	−0.50 (0.13)
Average $d = 0.28$ ($n = 24$)			

word type and visual presentation were significant ($F(2,46) = 10.8$, MSe $= 0.5$, $p < .001$, $F(1,23) = 20.2$, MSe $= 0.99$, $p < .001$, respectively). The interaction of word type and visual presentation was significant ($F(2,46) = 3.30$, MSe $= 0.15$, $p < .04$). A Tukey post-hoc analysis revealed that the differences in bias between either type of words and between the non-words were significant ($p < .05$). The greater bias for a "no" response found for non-words, in the non-matching condition relative to the matching condition, was significant as well. The difference in bias between the high- and the low-frequency words was not statistically reliable.

The results of Experiment 2 suggest that the addition of vowel marks did not produce an effect of bias to say "yes" in the matching condition for non-words. It could be pointed out that there was a significant greater tendency to say "no" when non-words appeared in the non-matching condition relative to the matching condition. However, even if the absolute relative difference between the two visual conditions serves as a measure for the effect, this difference was almost twice as large for words than for non-words, as revealed by the significant two-way interaction. Thus, although the vowel marks conveyed an unequivocal phonemic structure for the printed non-words, and allowed the generation of a phonological representation for both words and non-words, the difference in bias between words and non-words remained unchanged. This suggests that the phonetic representation that includes the amplitude envelope information was available only for words to influence the subjects' judgment. The overall similarity in the effects of bias in Experiments 1 and 2 is striking. This outcome confirms that the bias is independent of the print spelling-to-sound regularity, and provides additional support for the claim that the effect is lexically mediated.

General Discussion

The present study investigated the source of readers' ability to detect a correspondence between a printed word and its amplitude envelope. Experiment 1 revealed that matching print caused a bias to detect speech in a noise amplitude envelope, even in the unvoweled Hebrew orthography. This effect of bias was demonstrated only for words, whether high- or low-frequency, and not for non-words. In Experiment 2 we found an identical pattern of bias when the printed words were voweled, and therefore were phonologically unequivocal. All voweled words produced the effect, but not the voweled non-words. Moreover. the overall difference between the matching and the non-matching conditions was much larger for words than for non-words.

The bias to perceive speech embedded in amplitude-modulated noise derives from an automatic detection of correspondence between the printed letter string and the speech envelope related to it. The present study was concerned with how exactly this correspondence is detected. In order to match the visual to the auditory information, subjects had to generate from the print the relevant amplitude envelope. We examined whether this can be done by simply applying spelling-to-sound conversion rules to assemble a phonological representation, and by generating the envelope on-line from a phonetic structure that is contingent on the phonological representation derived from the print. The results of both experiments suggest that it cannot. Subjects did not show any bias to detect speech in the noise in the matching condition when non-words were presented. Note that Frost et al. (1988) did find a small effect of bias for non-words in the matching condition. This small effect for non-words is reflected in the present study by the greater tendency to say "no" in the non-matching condition relative to the matching condition. This tendency might be related to the overall lower detectability level obtained in the present study. In any event, the absolute difference in bias in the matching relative to the non-matching condition was much larger for words than for non-words.

Although a phonetic representation could have been easily generated from the printed non-words when they were voweled, the difference in bias between words and non-words remained unchanged. This outcome suggests that the bias effect is independent of spelling-to-sound regularity. Moreover, the effect seemed unaffected by the speed of print processing. Since the phonetic representation of low-frequency words is slower to generate from the print, the strong bias effect found for low-frequency words relative to high-frequency words suggests that speed of print processing is not a crucial determinant of the effect. Note that the addition of vowels in Hebrew was previously shown to accelerate the phonological processing of low-frequency words more than for high-frequency words (Koriat, 1984). Nevertheless, the bias effect found for low-frequency words did not increase in the voweled condition relatively to the unvoweled condition.

The stronger bias obtained for the low-frequency words might be related to the phonetic features of the stimuli employed. The magnitude of the bias effect depends among other things on the distinctiveness (or uniqueness) of the amplitude envelope, which affects the clarity of correspondence between the amplitude envelopes presented auditorily, and the word depicted by the print. It is possible that for some low-frequency words this correspondence was exceptionally clear. Recent results by Frost (submitted) support the conclusion that the bias effect is not affected by word frequency per se. In this study the bias for high- and low-frequency phonological alternatives of heterophonic homographs was examined, with an identical signal-to-noise ratio. The results demonstrated very similar bias effects for the high- and the low-frequency phonological alternatives (0.55 and 0.51, respectively).

Taken together, the results of Experiments 1 and 2 suggest that the effect of bias reported in the present and in previous studies is lexically mediated. We assume that the printed word addressed a lexical entry which contained, among other phonological and phonetic information, the word's amplitude envelope. Thus, the envelope was retrieved from the mental lexicon. By this view, a strong effect of bias can be shown only if the printed letter string can be related to an existing lexical entry. Non-words do not satisfy this requirement, and therefore did not produce the effect to the same extent.

The conclusion that envelopes are stored as lexical representations is supported by a recent study that examined the influence of lipreading on detection of speech in noise (Frost, Repp, & Zsiga, submitted). This study examined the effect of a visual presentation of a speaker's face on the detection of words and non-words in amplitude-modulated noise. The results demonstrated that an audio-visual match created a strong bias to respond "yes" when the stimuli were words, whereas no bias emerged when the stimuli were non-words. In contrast to orthographic information, there is a natural isomorphism between some visible articulatory movements and some acoustic properties of speech. Thus, the relations of articulatory movements to phonological and phonetic structure are non-arbitrary, and the correspondence between articulatory information and amplitude envelopes may be perceived without lexical mediation. Nevertheless, subjects did not produce any bias in the matching condition for non-words.

The proposal that amplitude envelopes are contained as holistic acoustic patterns in the mental lexicon is consistent with a view that lexical representations of spoken or printed words are not exclusively phonological. Models of speech perception often assume that the speech-processing system transforms the physical acoustic pattern into a more abstract linguistic representation which makes contact with the lexicon during word recognition. Regardless of the nature of this representation (i.e., what specific unit serves for activating a lexical candidate), lexical access is often viewed as a process which mediates access to more abstract linguistic information (e.g., Mehler, 1981; Pisoni & Luce, 1987). Our present results seem to suggest that the information contained in the lexicon is richer. In the present study, the presentation of a printed word resulted in the retrieval of an acoustic template – the word's amplitude envelope – from the lexicon.

At first glance, storing the word's amplitude envelope as a holistic pattern might seem to be without apparent benefits. The envelopes cannot identify a specific lexical candidate. However, they do convey prosodic and segmental information (e.g., speech timing, number of syllables, relative stress, and several major classes of consonant manner), that might help in selecting a lexical candidate among a highly constrained set of response alternatives (Van Tasell et al., 1987). Thus, the amplitude envelope might serve as additional information used by the listener in order to identify spoken words which have several acoustic realizations, or whose phonemic structure was not clearly conveyed (cf. Gordon, 1988). In these cases, a match between the perceived amplitude envelope and the stored template might confirm the identity of a lexical candidate. Clearly, richer representations do not constitute a parsimonious storage system. Nevertheless, the advantage of a more complex representational system is that it often allows a more efficient performance of the native speaker/listener.

One possible role of amplitude envelopes can be suggested in regard to the psychological distinction between words and non-words. It is often assumed that positive lexical decisions given to a letter string or to a spoken phonemic sequence are based on their relation to a semantic representation, whereas negative decisions result from the lack of such connections to the semantic network. In other words, positive and negative decisions are related to the meaningfulness of the presented stimuli. The results of the present study suggest possibly a different type of criterion. If words address stored amplitude envelopes and non-words do not, fast lexical decisions might be based, at least in part, on whether the printed letter string invoked a detailed phonetic representation such as the amplitude envelope. According to this interpretation, one factor that differentiates between words and non-words, and contributes to the word/non-word differences in the lexical decision task, is the generation of a phonetic code that contains envelope information. This suggestion, however, remains speculative and deserves further investigation.

The present study has additional relevance to old and recent debates concerning the processing of printed words. Models of printed word recognition are in disagreement concerning the extent of phonological recoding during visual word recognition. One important controversy relates to the automaticity of phonological recoding. It is often assumed that phonological recoding is very slow to develop, and lexical access occurs (with the possible exception of very infrequent words) directly from the visual structure of the printed words to meaning. This view is supported by results demonstrating that spelling-to-sound regularity affects lexical decisions only for low-frequency words (e.g., Seidenberg et al., 1984). In contrast, several studies have suggested that phonological information is available very rapidly as part of visual access to the lexicon (Perfetti et al., 1988; Van Orden et al., 1988). Perfetti and his colleagues have shown that the effect of a pseudo-word mask on the perception of a target word was reduced if there was a phonemic similarity between mask and target (i.e., "made", "mayd"). Using a different experimental technique, Van Orden (1987) and Van Orden et al. (1988) showed that when subjects had to decide whether a visually presented word belonged to a semantic category, they often made errors to homophones or pseudohomophones of category instances (i.e., positive responses were given to "rows" in the category of flowers, or "sute" in the category of clothing). These results were taken to demonstrate that phonological recoding occurs automatically and pre-lexically during lexical access.

The results of the present study support the view that phonetic recoding occurs automatically following the presentation of a printed word. What our results teach us is that the processing of a printed word results not only in a prelexical phonological representation but also in a very detailed phonetic speech representation, which is lexical and includes the word's amplitude envelope. This representation is *automatically* retrieved from the lexicon. Note that in this and previous studies which used the speech detection technique, subjects were not required to respond to the printed information. Nevertheless, they detected automatically the correspondence between the visual stimulus and the speech envelope.

In summary, the present study suggests that the presentation of words in the visual or the auditory modalities results in the generation of a rich array of orthographic, phonological and phonetic representations. One of these representations is the word's amplitude envelope. Because each of these representations may contact the mental lexicon, auditory illusions can be caused by visual printed information. The bias to detect speech in noise is caused by matching print because printed information arouses very detailed speech codes.

References

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographic and phonemic coding for lexical access: Evidence from Hebrew. *Journal of Experimental Psychology: Learning Memory and Cognition*, *10*, 353–368.

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

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

Frauenfelder, U.H., & Tyler, L.M. (1987). The process of spoken word recognition: An introduction. *Cognition*, *25*, 1–20.

Frost, R. (in press). Orthography and phonology: The psychological reality of orthographic depth. In: M. Noonan, P. Downing, & S. Lima (Eds.), *Literacy and Linguistics*. Benjamins.

Frost, R. Phonetic recoding of phonologically ambiguous printed words: Evidence from a speech detection task. Manuscript submitted for publication.

Frost, R., & Katz, L. (1989). Orthographic depth and the interaction of visual and auditory processing in word recognition. *Memory and Cognition*, *17*, 302–311.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 104–115.

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

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.

Gordon, P.C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception and Psycholophysics*, *43*, 137–146.

Hanson, V.L. (1981). Processing of written and spoken words: Evidence for common coding. *Memory and Cognition*, *9*, 93–100.

Hillinger, M.L. (1980). Priming effects with phonemically similar words: The encoding bias hypothesis reconsidered. *Memory and Cognition*, *8*, 115–223.

Jakimik, J., Cole, R.A., & Rudnicky, A.I. (1980). *The influence of spelling on speech perception*. Paper presented at the XXI annual meeting of the Psychonomic Society, St. Louis, Missouri.

Jusczyk, P.W. (1985). Auditory versus phonetic coding of speech signals during infancy. In: J. Mehler, M. Garrett, & E. Walker (Eds.), *Perspectives in mental representation: Experimental and theoretical studies of cognitive processes and capacities*. Hillsdale, NJ: Lawrence Erlbaum.

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

Klatt, D.H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. In: R.A. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Lawrence Erlbaum.

Klatt, D.H. (1989). Review of selected models of speech perception. In: W. Marslen-Wilson (Ed.), *Lexical representation and process*. Cambridge, MA: MIT Press.

Koriat, A. (1985). Lexical access for low- and for high-frequency words in Hebrew. *Memory and Cognition*, *13*, 37–44.

Layer, J.K., Pastore, R.E., & Rettberg, E. (1990). *The influence of orthographic information on the identification of an auditory speech event*. Paper presented at the 119th annual meeting of the Acoustical Society of America, Pennsylvania.

Luce, R.D. (1963). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society*, *B295*, 333–352.

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

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

Repp, B.H., Frost, R., & Zsiga, E. (1991). Lexical mediation between sight and sound in speech reading. Manuscript submitted for publication.

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

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

Seidenberg, M.S., Waters, G.S., Barnes, M., & Tanenhaus, M.K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior*, *23*, 383–404.

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

Van Orden, G.C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition*, *15*, 181–198.

Van Orden, G.C., Johnston, J.C., & Halle, B.L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 371–386.

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.

Van Tassel, D.J., Soli, S.D., Kirby, V.M., & Widin, G.P. (1987). Speech waveform envelope 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.

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.