Journal of Experimental Psychology: General
1981, Vol. 110, No. 4, 474-494

0096-3445/81/1004-0474$00.75

Phonemic Restoration: Insights From a New Methodology

Arthur G. Samuel
University of California, San Diego

SUMMARY

Phonemic restoration is a powerful auditory illusion in which listeners "hear" parts of words that are not really there. In earlier studies of the illusion, segments of words (phonemes) were replaced by an extraneous sound; listeners were asked whether anything was missing and where the extraneous noise had occurred. Most listeners reported that the utterance was intact and mislocalized the noise, suggesting that they had restored the missing phoneme.

In the present study, a second type of stimulus was also presented: items in which the extraneous sound was merely superimposed on the critical phoneme. On each trial, listeners were asked to report whether they thought a stimulus utterance was intact (noise superimposed) or not (noise replacing). Since this procedure yields both a miss rate *P*(intact|replaced), and a false alarm rate *P*(replaced|intact), signal detection parameters of discriminability and bias can be calculated. The discriminability parameter reflects how similar the two types of stimuli sound; perceptual restoration of replaced items should make them sound intact, producing low discriminability scores. The bias parameter measures the tendency of listeners to report utterances as intact; it reflects postperceptual decision processes.

This improved methodology was used to test the hypothesis that restoration (and more generally, speech perception) depends upon the bottom-up confirmation of expectations generated at higher levels. Perceptual restoration varied greatly with the phone class of the replaced segment and its acoustic similarity to the replacement sound, supporting a bottom-up component to the illusion. Increasing listeners' expectations of a phoneme increased perceptual restoration: missing segments in words were better restored than corresponding pieces in phonologically legal pseudowords; priming the words produced even more restoration. In contrast, sentential context affected the postperceptual decision stage, biasing listeners to report utterances as intact. A limited interactive model of speech perception, with both bottom-up and top-down components, is used to explain the results.

Phonemic restoration is a powerful auditory illusion in which listeners "hear" parts of words that are not really there. In the first report of the illusion, Warren (1970) replaced the first /s/ in "legislatures" with a cough or a tone in the sentence, "The state governors met with their respective legislatures convening in the capital city." Listeners were given typewritten versions of the sentence and asked to circle the location of the cough or tone and say whether it had replaced a speech sound. Localization of the extraneous sound was poor, and essentially all of the subjects reported that the sentence was intact; they apparently restored the deleted /s/.

In the decade since this experiment was reported, phonemic restoration has been very widely cited and very little studied. Only a handful of studies of the illusion have appeared in the literature. The most thorough of these was done by Warren and Obusek (1971). Using the same sentence as Warren (1970), Warren and Obusek varied the nature of the replacement sound, the information given to the subjects, and the size of the replaced chunk of speech. As in the first study, two measures of restoration were used: the distance metric (how accurately subjects located the replacement sound) and the hit rate (how often subjects accurately detected the deletion). Using these measures, the authors concluded that coughs, tones, and buzzes were essentially equally effective in producing restoration, and all were better than leaving a silent gap where the phoneme had been. The null difference among the different replacement sounds may have been due to a ceiling effect. Using similar methodology, Layton (1975) found that all replacement sounds are not equal; a tone is less effective than noise in producing restoration. Warren and Obusek's 8-dB amplitude manipulation of the replacement sound was ineffective, but even the quieter version was as loud as the sentence peak level, making strong conclusions unwise. The two most surprising findings were that the whole syllable "gis" was about as restorable as the "s" alone, and that informing subjects beforehand that something was in fact deleted did not help—restoration was still quite strong.

Obusek and Warren (1973) used a different approach to the problem of understanding restoration. They took a single word, "magistrate," and created several versions of it, replacing the /s/ with noise, the /gis/ with noise, or the /s/ with silence. Each version of "magistrate" was made into a tape loop and presented to listeners. The rationale was clever: It was already known that listeners hear (synthesize) very different words when a word is played over and over; this is the "verbal transformation effect" (Warren, 1961). Obusek and Warren reasoned that if restoration and the verbal transformation effect are both due to the listeners' synthesis of speech, then the transformations with the restoration stimuli should be enhanced at the point of replacement (/s/ or /gis/). In fact, they obtained exactly this result.

The only other published study of restoration was done by Warren and Sherman (1974). Three methodological improvements were made in this study. First, the splicing was done electronically, allowing greater accuracy. Second, seven sentences (rather than the single "legislatures" sentence) were used, a step toward generality. Third, in the original recording of the stimuli, the critical phonemes were mispronounced to insure that no appropriate cues remained. The theoretical issue considered was whether *subsequent* context could influence what phoneme was restored when more than one fit the prior context. Unfortunately, the operational definition of subsequent context, the remainder of the word, was a very limited one. For example, a sentence might be contextually neutral up to "deli* . . .", where * is the replacement and ". . ." is either "ery" or "eration," producing either "delivery" or "deliberation." Not surprisingly, this "subsequent context" was effective in influencing whether /b/ or /v/ was restored.

Despite the methodological improvements of Warren and Sherman (1974), there are several problems inherent in the way phonemic restoration has been studied. In particular, the measures of restoration have serious shortcomings. The distance measure (localizing the replacement sound) is, as Warren and Obusek (1971) admitted, at best an indirect measure of restoration. It is not at all clear why a subject who mislocalizes the noise by six phonemes should be considered to have restored more than one who misses by only three phonemes; both have restored the critical phoneme or been confused by the apparent location of the noise. The hit rate measure of restoration is considerably better in that it presumably is directly related to the restoration phenomenon. However, it has a serious drawback in that there is no way to obtain a false alarm rate (reporting that something was missing when nothing was), since the noise always replaces a phoneme. Thus there is no way

This research was conducted at the University of California, San Diego, and was supported by U.S. Public Health Service Grant MH 15828 to the Center for Human Information Processing, University of California, San Diego, and by National Science Foundation Grant BNS76-15024 A01 to Dave Rumelhart. Bell Laboratories provided support for the preparation of the manuscript.

I would like to thank those who have helped me in conducting this project. First and foremost, I thank Dave Rumelhart, who has provided invaluable advice and support throughout the project. I have also benefited from the suggestions of Donna Kat, Mark Liberman, Jay McClelland, Don Norman, and the LNR Research Group.

Requests for reprints should be sent to Arthur G. Samuel, who is now at the Department of Psychology, Box 11A Yale Station, Yale University, New Haven, Connecticut 06520.

to know whether a hit (correct report of something missing) is due to a failure to restore or to some sort of bias. An extreme case of bias would be a subject who never said that a phoneme was missing. In the studies reviewed here, that subject would have to be considered a perfect restorer. With no false alarm data, restoration is fundamentally indistinguishable from bias.

Clearly, other means are needed to measure phonemic restoration. The approach taken in the present study is to structure the task in such a way that the false alarm rate can be determined, thus providing a means to factor out bias. To do this, a second type of stimulus is used in addition to the normal restoration item. In the new stimulus type, the "replacement" sound is merely *added to* the appropriate phoneme, rather than replacing it. Stimuli thus come in pairs; in one version, a sound replaces a phoneme, whereas in the other the same sound coincides with the phoneme. The rationale for this approach is simple. The phenomenology of phonemic restoration is that the stimulus word seems to be intact and an additional sound appears to be overlaid over part of it. The new version of the stimulus corresponds exactly to this description. Given appropriate controls (for such factors as masking), to the extent that the two versions sound alike, restoration is indicated. Since a false alarm rate is available (probability of reporting "replaced" when presented with an added item), bias can be factored out using signal detection theory.

Using this improved methodology, the experiments reported here address a fundamental question: How does lexical access occur from speech? In answering this question, the focus will be on determining how much the system depends on analyzing the acoustic information into words and how much it uses higher level knowledge to derive lower level information; in short, what are the relative contributions of bottom-up and top-down analysis? The metatheoretical framework that prompted this approach is the interactive schema theory of Rumelhart (1977). In this model, the perceptual–cognitive system is viewed as a collection of active processing units (schemata). The schemata are data structures that are activated either by other schemata (top-down) or by incoming sensory information (bottom-up). In Rumelhart's model, the perceptual process is essentially an evidence-gathering task; perception occurs when a given schema has received sufficient evidence. The pieces of evidence can come from both bottom-up and top-down sources. In general, the top-down information can be characterized as *expectations*, and the bottom-up as *confirmation*. The restoration phenomenon itself is clear evidence for top-down processing of speech; the listener uses the linguistic context to restore the appropriate (expected) phoneme. However, the illusion also depends on bottom-up factors; the amplitude, spectrum, and very presence of the replacement sound can determine the success of restoration. Thus study of a product of the speech recognition process, restoration, may provide insights into the general working of the system.

Within this framework, a number of processing issues are considered. In the first experiment several factors that could affect top-down and/or bottom-up analysis of speech are manipulated: word frequency and length, and phone class and location within the word of the replaced sound. These factors are examined using words produced in isolation, in order to factor out the influence of variables at levels higher than the lexical. In the second experiment, a priming paradigm is used to investigate the role of lexical and phonological knowledge in the restoration phenomenon. The final experiment examines the effects of memory load and sentential context on phonemic restoration.

Experiment 1

The basic hypothesis of the present study is that phonemes are restored on the basis of expectations and confirmation; the stronger the contextual constraints, the greater the expectations should be, and the less bottom-up confirmation that should be needed.

Four factors are experimentally manipulated in order to vary the level of expectation and confirmation of a particular phoneme. First, high-frequency words are compared with low-frequency words. High-frequency words are more easily retrieved from the lexicon. To the extent that lexical information provides expectations of constituent phonemes, more restoration should occur in more frequent words. Second, two-syllable, three-syllable, and four-syllable words are compared. To the extent that a longer word provides more context, restoration should be enhanced. Third, five different classes of phonemes are investigated: liquids, stops, nasals, fricatives, and vowels. The replacement–added sound in the experiment is a burst of white noise. The degree of similarity between this noise and the five phone classes produces a variation in bottom-up confirmation; the fricatives (and to a lesser extent the stops) should be most easily restored when white noise is the inserted sound. The final factor is the position within a word of the replacement–addition: word-initial, word-medial, and word-final phonemes are manipulated. Several predictions can be tested for this factor. If expectations can be generated on-line and very quickly, then word-final position should exhibit the most restoration, since the earlier parts of the word serve as context. If simple forward or backward masking is involved, word-medial position should be most susceptible. Finally, several theorists have recently argued that the initial syllable is disproportionately important in lexical access (Cole, 1973; Foss & Blank, 1980; Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978; Taft, 1979). If this is so, we may expect listeners to be particularly sensitive to any changes made in word-initial position.

Method

Stimuli. Each test word was spoken clearly by the author and recorded on audiotape. The word was then amplified, low pass filtered (5 kHz), digitized (12-bit A/D at 10 kHz sample rate), and stored on a PDP-11 disc file. Using a waveform editor, two versions of the word were constructed from the stored file, a *replacement* item and an *added* item. Figure 1 presents oscillograms of the critical segment of the original, added, and replaced versions of the word "funerals." In this example, the medial /n/ was the target phoneme. To create the test items, the /n/ was located visually on an oscilloscope and auditorily over headphones. Once its general location was determined, the segment was bracketed by two pointers according to the following criteria: (a) when the part of the word *before* the pointer to the segment onset was played, no trace of the target phone was audible; (b) when the part of the word *after* the pointer to the segment offset was played, no trace of the target phone was audible; (c) on the oscilloscope, all parts of the pattern characteristic of the target phone were included between the onset and offset pointers. As is evident in Figure 1, satisfying (a) and (b) generally insured (c). In fact, due to coarticulation, substantial portions of the adjacent phonemes were virtually always deleted to meet these criteria. Using this procedure produced critical segments 50–279 msec long.

Once the target phone was properly bracketed, the digital root mean square amplitude (DRMSA) of its central 20 msec was computed. To create the added version, a random number was added to each point within the target segment (this is the digital equivalent of mixing in white noise). Each random number was between +DRMSA and −DRMSA. Thus the amplitude of the noise added to (or replacing) the target was a function of the amplitude of the target itself. Using this procedure insured that the S/N ratio would be constant for all targets, regardless of the original amplitude of the target. The panel labeled ADDED in Figure 1 shows the results of this procedure. The added version of the word was stored on disc file. The replaced version was constructed by simply replacing each point within the bracketed segment with a random number between +DRMSA and −DRMSA.[1] The REPLACED panel in Figure 1 illustrates the results. The replaced item was also stored on disc file.

Replaced and added versions of 90 test words were constructed. Four factors were represented in these 90 words:

1. Word frequency: 45 high-frequency words (100–300 occurrences/million) and 45 low (1 occurrence/million, Kucera & Francis, 1967). The low-frequency words were all recognizable English words.
2. Word length: 30 two-syllable, 30 three-syllable, and 30 four-syllable words.
3. Phone class of the replaced/added phoneme: 18 liquids, 18 stops, 18 vowels, 18 fricatives, and 18 nasals.
4. Phone position: 30 word initial, 30 word medial, and 30 word final. The 90 words thus represent a 2 × 3 × 5 × 3 factorial crossing of these four factors (Frequency × Length × Class × Position).

Matched to the 180 test-word items (90 added and 90 replaced) were 180 *control segments.* Each control segment was the segment of the test word that contained white noise. Thus, for replacement items, the control segment was the segment of white noise that had replaced a phone; for added items, the control segment was a mixture of speech and noise.

The control segments were used to determine the basic discriminability of the added and replaced versions of the test words. Since restoration will be measured by the difficulty of discriminating added and replaced test words, it is important to obtain a measure of the discriminability of the critical segments outside of their

[1] The amplitude of the replacement sounds in the present study was generally lower than that of replacement sounds in previous studies by Warren (e.g., 1970) and his colleagues. This may reduce the amount of restoration observed.

linguistic context. With other appropriate controls, good discriminability of the control segments, along with poor discriminability of the test words, is evidence for restoration.

Table 1 presents a summary of some of the important acoustic characteristics of the test words and control segments.

Apparatus. All stimuli were stored on disc file in a PDP-11/45 computer. For presentation to the subjects, they were output through a 12-bit D/A converter, amplified, low pass filtered (5 kHz), and played binaurally over stereo headphones. Subjects heard the stimuli in individual acoustically shielded booths and responded by pressing one of two labeled buttons, using the first two fingers of the right hand. All experimental events were controlled by the PDP-11.

Procedure. On each trial, subjects heard one test item (a word or a segment). For words, they were told to push one button if the noise replaced part of the word, and the other if it coincided with part of the word. For segments, they pushed one button if they heard just white noise, and the other if the white noise was mixed with another (speech) sound. The nature of the stimuli was fully explained beforehand. Subjects were told to respond as quickly as possible without sacrificing accuracy. They were instructed to guess when necessary; a response was required on each trial. Groups of 1-4 subjects were tested simultaneously. When all subjects had responded (or 5 sec had elapsed), the computer waited 1 sec, and then initiated the next trial.

Words and segments were presented in blocks. Twenty practice words (half added, half replaced) and twenty practice segments (half noise, half noise plus speech) were presented first. The 90 test words were then presented in a random order. On each trial, it was randomly determined whether the added or replaced version of the word would be presented. After the word block finished, a block of 90 segments was run. These segments were taken directly from the test words that had just been presented; if the sixth test word was the replaced version of "modern," the sixth segment would be the noise burst that had replaced a phoneme in "modern." The block of segments was followed by another block of 90 words and a block of 90 segments containing the versions of the words and segments not presented in the first pass. The design was thus within subject; each subject received both versions of each word and the control segments from each version.

All trials were run in a single session that lasted approximately 20 min.

Subjects. Twenty subjects participated in Experiment 1. All were native English speakers with no known hearing problems. They received course credit for their participation.

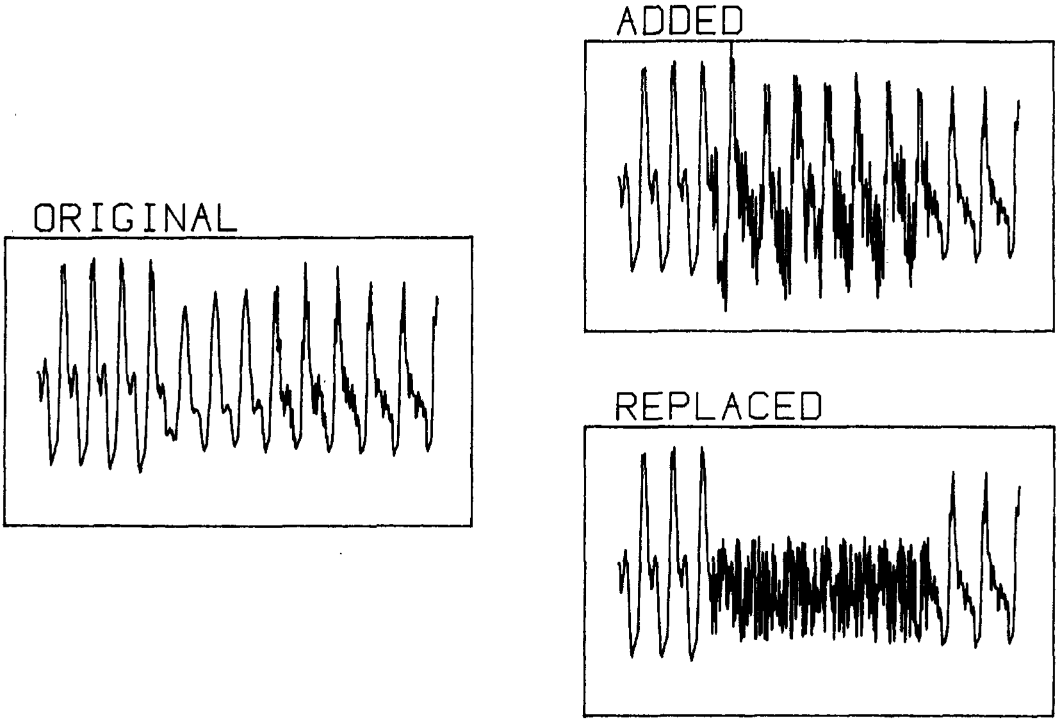

Figure 1. Oscillograms of a part of the word "funerals." The left panel is from the original digitized version of the word, and includes the /n/ and part of the preceding /u/ and succeeding/ə/. (The ADDED panel is the same part of the word, with digital white noise superimposed on the critical segment. The REPLACED panel is the version of the word with digital white noise replacing the critical segment.)

Results

Three related measures of restoration were calculated from the data. The *miss rate* is the probability that a replacement item was labeled *added*(or, for the segments, that a noise burst was labeled *noise plus speech*). This measure is essentially the same as the accuracy measure used in previous studies. However, it is more valid in this study because subjects were presented with items that really were intact. The primary measure of restoration is d', the discriminability of added and replaced versions of the same words. To the extent that d' is near zero, the two versions are not discriminable; the inference is that they sound alike because the missing phoneme is being restored in the replaced version. This inference is rendered considerably more plausible if the d' for the corresponding segments is far from zero. A high d' for segments indicates that in isolation, the critical segments do not sound alike. In particular, it rules out a simple masking explanation for the low d' for words.[2] The third measure, Beta, is an index of bias. A Beta greater than 1.0 reflects a tendency to report that the word was intact, regardless of whether the item was a replaced or added version; values less than 1.0 mean the reverse (for segments, a Beta greater than 1.0 means a bias toward saying "noise plus speech"). High Beta values reflect a form of response bias in the postperceptual decision process.

Table 1
Stimuli Analysis for Words and Segments in Experiment 1

Stimulus type	Word duration	Segment duration	Segment amplitude
Frequency			
High	660	144	745
Low	651	133	597
Length			
Two-syllable	585	140	693
Three-syllable	655	141	557
Four-syllable	725	135	763
Phone class			
Liquid	624	127	779
Stop	707	114	610
Nasal	649	127	697
Fricative	717	193	268
Vowel	580	131	1000
Position			
Initial	668	115	825
Medial	666	107	819
Final	633	193	369

Note. The word and segment durations are in milliseconds. The segment amplitudes are the means of the digital RMS amplitudes (DRMSA), expressed in arbitrary units. These values are scaled such that vowels have a mean of 1000; the numbers should be treated as relative values.

Table 2
Miss, d′, *and Beta for Words in Experiment 1*

Stimulus type	Miss	d'	Beta
Frequency			
High	44	.88	1.28
Low	41	1.06	1.36
Length			
Two-syllable	34	1.10	1.15
Three-syllable	43	1.02	1.39
Four-syllable	50	.77	1.37
Phone class			
Liquid	34	1.47	1.57
Stop	60	.36	1.17
Nasal	40	1.13	1.43
Fricative	63	.36	1.20
Vowel	15	1.69	.72
Position			
Initial	30	1.00	.98
Medial	56	.89	1.71
Final	41	1.08	1.37

Note. The miss rates are percentages; d' and Beta are in absolute units.

Table 2 presents the miss, d', and Beta scores for the words, broken down by word frequency, word length, phone class, and phone position. The scores for each factor were derived by pooling all of the data from all of the subjects, and collapsing across all other factors (this was necessary to assure an adequate number of observations in each cell). A matrix of the same form as Table 2 was also calculated for each subject. Using these scores, 12 one-way analyses of variance were conducted, 4 factors (frequency, length, class, and position) × 3 measures (miss, d', and Beta).[3]

[2] An explanation based on forward and backward masking is not ruled out by this control; it is considered further in Experiment 2.

[3] Separate one-way analyses of variance were conducted for each factor, rather than the usual multifactor

The effect of word length was as predicted; longer words provided more context and greater restoration. The large effect on the miss rate, $F(2, 38) = 19.06$, $p < .001$, is due to both perceptual, d', $F(2, 38) = 5.20$, $p = .01$, and postperceptual factors, Beta, $F(2, 38) = 6.88$, $p < .01$.

As can be seen in Table 2, the effect of frequency was less-clear-cut. An expectation-based model predicts more restoration of phonemes in high-frequency words than low and thus worse performance. In fact, the miss rate for high frequency words is slightly higher, but not significantly so, $F(1, 19) = 2.85$, $p > .10$. The frequency conditions also did not differ in bias, $F(1, 19) = 1.89$, $p > .10$. The d' measure does reach significance, $F(1, 19) = 7.21$, $p < .05$, suggesting that more frequent words, through more efficient lexical access, yield more restoration. However, the effect is not very robust; under similar conditions, Samuel (1981) found no difference on any of the three measures as a function of word frequency. Thus, acceptance of a frequency effect must be considered tentative, pending further data.

Even a cursory inspection of Table 2 reveals the massive effect of phone class on restoration. As predicted by simple acoustic factors, fricatives and stops are very difficult, whereas the more periodic phones are relatively easy. Recall that restoration is hypothesized to be due to the bottom-up confirmation of the listener's expectations. Since expectations of each phone class were presumably equivalent, the data suggest that a burst of white noise is more effective at confirming the presence of fricatives or stops than liquids, nasals, or vowels. The analyses of variance for the phone class factor bear this out: miss, $F(4, 76) = 51.97$, $p < .001$; d', $F(4, 76) = 43.60$, $p < .001$; Beta, $F(4, 76) = 3.50$, $p < .05$. The effect is thus primarily one of discriminability rather than bias.[4]

Table 3
Miss, d', *and Beta for Segments in Experiment 1*

Stimulus type	Miss	d'	Beta
Frequency			
High	6	2.71	.54
Low	9	2.60	.86
Length			
Two-syllable	7	2.83	.86
Three-syllable	8	2.52	.65
Four-syllable	7	2.61	.63
Phone Class			
Liquid	6	3.12	1.03
Stop	9	1.97	.48
Nasal	8	3.08	1.32
Fricative	6	2.55	.44
Vowel	8	3.16	1.82
Position			
Initial	10	2.54	.90
Medial	9	2.77	1.04
Final	4	2.80	.31

Note. The miss rates are percentages; d' and Beta are in absolute units.

An examination of the comparable data for the control segments (Table 3) aids in interpreting the phone class result. The overall level of performance for the segments is much higher than that for the words, indicating that in general the acoustic information needed for discrimination of added and replaced versions was present. A two-way analysis of variance (Words vs. Segments × Phone Class) confirms the superior discriminability of the segments, $F(1, 19) = 170.51$, $p < .001$. Taking this absolute difference in level into account, we may look at the *relative* levels of performance within a given factor. Looking at the d's for segments of the various phone classes, we find

analyses, in order to assure an adequate number of observations in each cell for each subject. The signal detection parameters d' and Beta are extremely nonlinear for miss and false alarm rates near 0% or 100%; such extreme rates were fairly common for individual subjects when the data were factorially divided, since very few observations were obtained in some cells (e.g., each cell of the Phone Class × Position interaction contained $90/[5 \times 3] = 6$ observations). In order to avoid the unmanageable noise such an analysis would produce, separate analyses for each factor were necessary. This necessarily implies the loss of any interactions that might be present. However, few of the factors would be expected to interact (perhaps position with the others), and no crossover interactions are predicted by any current theory. Thus the main effects are sufficient.

[4] Samuel (1981) has replicated this pattern using white noise as the replacement sound. When a 1000-Hz tone was used instead, vowels were slightly better restored than fricatives, the reversal predicted by the confirmation view. The overall level of restoration was, however, greater with white noise than with the pure tone, suggesting that spectral similarity is weighted more heavily than periodicity.

that stops and fricatives are relatively difficult even in isolation. This result is no doubt due to the fact that stops and especially fricatives share more acoustic properties with the noise than do the other phone classes; they are also of relatively low amplitude (see Table 1) and more difficult to discriminate.

Although the segment data provide a partial explanation for the pattern of results on the words, there apparently is another factor to consider. Performance on the periodic segments was perhaps 50% better than that for stop and fricative segments; the comparable differences for words are around 300%. The analysis of variance reveals that this interaction of lexical status and phone class is reliable, $F(4, 76) = 4.52, p < .01$. Differential restorability, because of the acoustic differences, could account for the additional spread in performance.

The analysis of performance as a function of phoneme position revealed several interesting results. Although discriminability did not vary reliably, $F(2, 38) = 1.33, p > .20$, the differences in miss rates for initial, medial, and final phonemes were significant, $F(2, 38) = 26.69, p < .001$. This pattern was due to a significant difference in bias across the three positions, $F(2, 38) = 4.99, p < .05$. The bias results suggest that subjects are generally inclined to report "intact" unless there is evidence to the contrary. If a noise burst occurs near the middle of a word, where forward and backward masking reduce perceptibility, subjects tend to report that nothing was removed. Given this overall bias toward reporting "intact," the Beta value for initial phonemes (.98 vs. 1.71 for medial and 1.37 for final) indicates that the presence of an initial noise burst (added or replacing) biases the decision process toward reporting that something was amiss. This pattern was predicted from the view that lexical access is dominated by word-initial phonemes; *any* disruption in initial position is particularly noticeable and therefore leads to a bias toward saying "replaced."

No support was found for a rapid on-line effect of lexical context. Such an effect would produce more restoration in final position, since earlier parts of the word would provide contextual cues. Neither the d′ nor the miss rate measure revealed this result. Apparently, in the isolated word situation tested here, within-word context cannot produce the expectation and confirmation required for restoration.

Discussion

The results from the first experiment provide support for the schema-based model outlined earlier. The fact that more restoration was found for longer words than for shorter words supports the view that restoration is a function of context; the greater the context, the greater the expectation, the greater the restoration. A small effect of word frequency was also obtained, providing some additional support for the role of expectations.

The importance of bottom-up confirmation was perhaps even greater than that of expectation in this single-word situation. A large effect of phone class was obtained, suggesting that the replacement sound must be compatible with the phoneme it is to restore. Even more basic acoustic factors, such as forward and backward masking, also appear to play some role. Thus, the pattern of results indicates that phonemic restoration occurs when the bottom-up signal (e.g., white noise) is sufficient to *confirm* the presence of a schema that is *expected* on the basis of context (e.g., the rest of the word). There is a trade-off between these two factors; when expectations are strong, less confirmation (i.e., an acoustic signal less like the actual phoneme) is needed for restoration, and vice versa.

Examination of the results for the various factors manipulated in Experiment 1 reveals how critical it is to factor changes in the simple miss rate into discriminability and bias components. Consider, for example, the phone class and position factors. If only miss rates were considered, one would be inclined to conclude that these two factors can have similar effects, for instance, that fricatives and medial phones both tend to be well restored. The signal detection analysis shows that these factors operate in different ways; the high miss rates for fricatives are in fact due to true perceptual restoration, whereas the high medial miss rates are the result of a postperceptual response bias.

This response bias can be difficult to interpret theoretically: What does it mean to be more biased toward reporting utterances as intact in some situations than in others? I believe its interpretation is made clearer by an introspective report. In many cases, the real answer to the question, "Was the utterance intact?" is "I don't know." That is, listeners often do not have a sufficiently detailed acoustic representation of the stimulus available to make the required judgment. Under these circumstances, bias effects will appear. The part of the cognitive system responsible for making the required choice uses all of the sources of information available, including questioning whether there was anything "wrong" with the stimulus. The answer is likely to be yes in initial position and no in medial. In contrast, the effect of phone class is truly perceptual and leads to some phonemes being restored better than others. Vowels, for example, are not restored very well, leading to an incomplete word percept; this percept is available for the required judgment and leads to an improvement in the d′ measure. As the results for the word length factor show, a particular manipulation can affect both true perceptual restoration and postperceptual bias. It is thus critical to be able to pull these effects apart to understand the restoration illusion.

Experiment 2

In the single word situation used in the first experiment, the kinds of knowledge that can be brought to bear are clearly more limited than in the usual discourse situation; syntax, semantics, and pragmatics are ruled out. However, listeners still have at least two powerful knowledge sources. First, they possess detailed (though probably implicit) knowledge of the allowable sequences of phonemes in English. Second, they have a rather detailed representation of the sound of each word in their memory. The second experiment is concerned with the role that these phonological and lexical knowledge sources play in the perception of speech.

One way to separate the contributions of these two factors is to use stimuli that follow the phonological rules of English but have no lexical entry in memory—pseudowords. The intuitive prediction is that if the same sort of restoration task is run with pseudowords and words, performance will be better on words than on pseudowords because of the familiarity of the words; the words should be easier to encode and examine, permitting better discriminability of added and replaced items. However, the top-down component of the schema model predicts the reverse; *because* the words are more familiar (i.e., have a lexical entry), listeners will be better able to generate expectations of deleted phonemes and should therefore restore more, producing poorer discriminability of the two versions.

There is a serious problem in running the experiment as it has just been outlined. The pseudoword condition is at a fundamental disadvantage because the words can be done by process of elimination, whereas pseudowords cannot. For example, consider the word item "basis" in which the final "s" has been replaced. If no restoration occurs, the listener hears "basi*" and can answer "replaced" because "basi*" is not a word, even if he or she thought the noise burst was on top of another phoneme. In the comparable pseudoword condition, "pafis" might be presented with the "s" replaced, yielding "pafi*." If the noise is mislocalized (as in the word case), the listener should report "added." Thus, because no lexical backup strategy is available, an advantage exists for the word condition if the experiment is run this way.

To overcome this problem, a cuing paradigm is used in Experiment 2. In this paradigm, the intact version (neither added nor replaced) of a test item is presented first, followed by the usual added or replaced version. For the pseudoword example given earlier, the trial would thus become "pafis" followed by "pafi*" or "pafi[*s]" (* = replacement sound, [*s] = replacement sound added to "s"). The listener thus knows what the original was and can use the same strategies in both conditions.

For the word items, the cuing paradigm can also be thought of as a priming condition, since the cue word should activate an entry in the lexicon. Again, the intuitive prediction clashes with an expectation-driven model's prediction. The simplest prediction would be that knowing what word was coming should help the subject to know what to listen for. The top-down alternative is that despite this advantage, performance should actually be worse because priming a word increases the expectation of each phoneme, increasing restoration.

Method

Stimuli and design. To address the various issues just discussed, three conditions were included in Experiment 2. Each condition was similar to Experiment 1 in format. In all of the conditions, the amplitude-matched white noise replacement procedure of Experiment 1 was used. All new stimuli were recorded by the same speaker under the same conditions as the words used in Experiment 1 and were constructed using the same digital editing techniques.

The *unprimed pseudoword* condition was identical to Experiment 1, except that each word was replaced by a matched pseudoword. Each pseudoword was constructed by changing one phoneme per syllable in the original word to a phonologically legal different phoneme from the same phone class (stops replaced stops, fricatives replaced fricatives, etc.). The critical phoneme was never changed, and was used as the added/replaced phoneme in the pseudoword as well. This procedure insured that the pseudowords were very closely yoked to the original words but were nevertheless nonlexical items. Each pseudoword was pronounced with the same stress pattern as the original word it replaced and of course had the same number of syllables. Some examples of word-pseudoword pairs (with the critical phonemes capitalized) are "prOgress-crOgless," "basiS-pafiS," "Mildew-Molbew," and "acTivity-ecTathigy."

Table 4 presents an analysis of the pseudowords and their control segments analogous to the breakdown of the words in Table 1. The similarity between the words and pseudowords is evident in the values in the two tables. The only consistent difference is a 10% lengthening of the pseudowords, probably produced by their unfamiliarity.

The unprimed pseudoword condition was included to determine how well listeners can do the added/replaced task in a situation that calls for actually determining whether the noise is superimposed on a phoneme. If this proves very difficult, it suggests that the primary strategy used on the words was the "failure to restore = nonword = replaced" strategy discussed earlier. Note that with this strategy, the methodology measures exactly what we would like: Failure to restore produces "replaced" responses, whereas "added" is reported when an intact word is heard (restored).

The same pseudowords were used in the *pseudoword/1500* condition, but on each trial, the intact version preceded the added or replaced test item. The time from onset of the cue pseudoword until onset of the test item was 1500 msec (since the cues were approximately 700 msec long, there was about 800 msec from cue offset to test item onset).

The *word/1500* condition was comparable to the pseudoword/1500 version, except that the test words from Experiment 1 were used (the stimuli were in fact identical). The comparison between performance in this condition and the pseudoword/1500 condition should clarify the role lexical knowledge plays in phonemic restoration. The comparison between performance in this condition and performance on the unprimed words of Experiment 1 should reveal the effect on restoration of priming a word.

Table 4
Stimuli Analysis for Pseudowords and Segments in Experiment 2

Stimulus type	Pseudoword duration	Segment duration	Segment amplitude
Frequency			
High	720	152	707
Low	724	145	660
Length			
Two-syllable	656	155	696
Three-syllable	717	149	652
Four-syllable	793	142	702
Phone class			
Liquid	713	136	716
Stop	757	144	569
Nasal	707	131	835
Fricative	771	195	298
Vowel	662	137	1000
Position			
Initial	726	124	802
Medial	742	123	895
Final	698	198	354

Note. The pseudoword and segment durations are in msec. The segment amplitudes are the means of the digital RMS amplitudes (DRMSA), expressed in arbitrary units. These values are scaled such that vowels have a mean of 1000; the numbers should be treated as relative values.

Apparatus and procedure. The apparatus was the same as that used in the previous experiment. The procedure was also the same except that in the cued conditions, subjects were told that they would hear a target first, followed by a test item. They were told to use the target to help decide whether the test item was an added or replaced item.

The segment controls followed the same pattern as the words (or pseudowords). For example, a replaced segment trial in the pseudoword/1500 condition would include the target phoneme (taken from the original intact pseudoword) followed by the noise burst that replaced that phoneme. The onset to onset time would be 1500 msec.

Sessions for the cued conditions lasted approximately 35 min.

Subjects. Sixty subjects participated in Experiment 2, 20 in each of the three conditions. All were native English speakers with no known hearing problems. They received $2 or course credit for their participation.

Results

Table 5 and Table 6 present the miss, *d'*, and Beta values for the words and pseudowords, and their control segments.

As might be expected, performance in the

Table 5
Miss, d', *and Beta for Words and Pseudowords in Experiment 2*

	Unprimed pseudowords			Pseudoword/1500			Word/1500		
Stimulus type	Miss	d'	Beta	Miss	d'	Beta	Miss	d'	Beta
Frequency									
High	54	.26	1.06	35	1.25	1.32	42	.79	1.17
Low	48	.43	1.06	37	1.30	1.48	48	.82	1.34
Length									
Two-syllable	49	.25	1.03	36	1.26	1.41	43	.92	1.28
Three-syllable	52	.44	1.13	33	1.36	1.37	47	.72	1.22
Four-syllable	52	.33	1.07	38	1.19	1.40	46	.78	1.24
Phone Class									
Liquid	50	.19	1.01	28	1.65	1.45	35	1.31	1.42
Stop	54	.31	1.09	42	1.05	1.39	64	.17	1.08
Nasal	52	.37	1.09	27	1.65	1.43	45	1.04	1.48
Fricative	58	.44	1.20	64	.51	1.38	61	.03	1.01
Vowel	41	.41	.99	17	1.65	.81	20	1.63	.96
Position									
Initial	47	.32	1.03	34	1.19	1.25	34	.94	1.06
Medial	61	.30	1.13	44	1.19	1.66	58	.83	1.68
Final	45	.42	1.03	29	1.45	1.29	43	.71	1.13

Note. The miss rates are percentages; d' and Beta are in absolute units.

unprimed pseudoword condition was extremely poor. Listeners apparently were unable to make an absolute added/replaced judgment. Rather, the strategy generally used was to determine whether the speech stimulus heard matched a representation in memory. If not, "replaced" was reported. Since this strategy was preempted in the unprimed pseudoword condition, performance was near chance.

The central question of Experiment 2 was whether lexical knowledge increases restoration. The answer seems to be yes; discriminability of added/replaced versions of primed pseudowords (pseudoword/1500) was 50% better than that for the comparable word condition. To test this difference statistically, a two-way analysis of variance was conducted using Word versus Pseudoword and Phone Class as the factors (the Phone Class factor was used to pull out variance from the effect of interest). The higher miss rates for words than pseudowords, $F(1, 38) = 8.05$, $p < .01$, was entirely due to perceptual restoration (d'), $F(1, 38) = 14.19$, $p < .001$; the postperceptual bias showed a nonsignificant trend in the opposite direction, $F(1, 38) = 3.14$, $p > .05$. Despite the large familiarity advantage that words enjoy, the availability of a lexical entry appears to increase expectations sufficiently to produce a discriminability disadvantage for the words.

One caveat needs to be added at this point. Examination of Table 6 reveals that performance on the primed pseudoword segments was also considerably better than that on the primed word segments; there was also an unusual pattern of bias effects. The reason for these differences is not clear—perhaps the slight lengthening of pseudowords (and thus their segments) relative to words enhanced performance. The segment controls from the unprimed pseudoword condition were intermediate in difficulty, easier than those for the words, but harder than those for the primed pseudowords, and the bias was in the normal range. In any event there is a large difficulty factor (due to familiarity) favoring the words over the pseudowords that does not operate in the segment conditions. In resource allocation terms, the pseudowords should require more resources to process, leaving less capacity for the discrimination judgment; this factor should have pushed performance on the words and pseudowords in the opposite direction from the results obtained. Thus it seems reasonable to tentatively accept the original conclusion that the existence of a lexical entry facilitates restoration and thus hurts discriminability. However, a replication of the pseudoword advantage is clearly needed before any strong claims can be made.

It seems reasonable that telling the subjects what word to listen for should be helpful, since the listener can presumably set up some sort of auditory image and see whether the test item matches it. However, in a system with expectation-driven processing, this strategy could backfire; the auditory image is in effect an expectation and thus should increase restoration, hurting performance. A comparison of discriminability in the word/1500 condition and the unprimed condition of Experiment 1 shows that performance is in fact *worse* when subjects know what word to expect. A more detailed comparison of the primed and unprimed conditions reveals that this degradation in performance varies with position of the critical phoneme. Recall that in several current theories (e.g., Cole, 1973; Foss & Blank, 1980; Marslen-Wilson & Welsh, 1978), lexical access is based on information in the first few phonemes of a word; these theories should predict more restoration (and thus poorer discriminability) for word-final phonemes than for initial ones, since the early part of a word provides context for later parts. For unprimed words, no such discriminability effect was observed. For primed words, however, the predicted pattern appears, with restoration increasing from initial through final position. In a two-way analysis of variance (Primed vs. Unprimed × Phone position), this interaction was marginally significant, $F(2, 76) = 2.65$, $.05 < p < .10$.

To see whether the degradation due to priming was reliable, the primed word condition was replicated. In the replication, the interval between cue *offset* and target *onset* was held at 200 msec; all other aspects of the experiment were unchanged. The d' values for initial, medial, and final position were 1.07, .84, and .78, respectively, replicating the Priming × Position interaction. In the replication, the interaction reached significance, $F(2, 76) = 4.06$, $p < .02$. Thus, the role of the beginnings of words appears to be confirmatory; in isolation, there are no expectations to confirm, whereas in the priming condition, listeners expect a particular word, and when their expectations are confirmed, the ending is restored.

Table 6
Miss, d′, and Beta for Segments in Experiment 2

	Unprimed pseudowords			Pseudoword/1500			Word/1500		
Stimulus type	Miss	d'	Beta	Miss	d'	Beta	Miss	d'	Beta
Frequency									
High	6	3.10	.91	6	3.67	2.74	8	2.45	.61
Low	7	2.78	.74	8	3.31	2.33	9	2.34	.67
Length									
Two-syllable	6	3.13	.98	6	3.88	4.54	7	2.72	.76
Three-syllable	8	2.85	1.04	9	3.28	2.06	9	2.29	.61
Four-syllable	5	2.85	.53	7	3.33	1.91	9	2.23	.59
Phone Class									
Liquid	6	3.45	1.96	3	4.21	2.58	5	3.14	.64
Stop	7	2.66	.67	5	3.34	1.14	9	2.10	.54
Nasal	8	3.16	1.51	1	4.55	1.28	6	3.05	.80
Fricative	3	2.73	.22	22	2.44	3.00	15	1.32	.61
Vowel	8	3.65	4.23	3	4.12	2.48	8	3.42	2.43
Position									
Initial	6	3.21	1.12	4	4.00	3.06	7	2.71	.66
Medial	10	2.63	1.07	7	3.44	2.28	8	2.61	.68
Final	3	3.12	.37	10	3.14	2.33	11	1.97	.62

Note. The miss rates are percentages; d' and Beta are in absolute units.

Discussion

The major finding of Experiment 2 is that lexical access from speech is both a top-down and bottom-up process, even in the relatively impoverished isolated-word situation tested. The availability of a lexical representation produced a 50% decrease in discriminability of replaced and added items, reflecting the role of lexical knowledge in the restoration phenomenon. Put another way, because of knowledge at the word level, the perception of the sounds of speech was changed.

This perception was also affected by cuing the subjects as to what word to expect. However, the effect of such cuing was to hurt performance, rather than its usual effect of helping. This reversal is predicted by a top-down model of speech perception, since the cue served to prime a lexical unit, increasing the expectation of its component sounds.

It is important to note that these effects are *perceptual* effects; they appear in the listener's ability to discriminate intact from interrupted words or pseudowords. The postperceptual bias measure was constant across the conditions under study, indicating that lexicality and priming have little effect on the decision stage. Thus, the data support a truly top-down effect, in which higher-level knowledge affects the operation of a lower level process.

Experiment 3

In the first two experiments, the added/replaced methodology illustrated the role that acoustic, phonological, and lexical factors play in the perception of isolated words. In the final experiment, this methodology is extended to investigate phonemic restoration (and thus speech processing) in a more interesting domain—sentences.

The general question being addressed is how higher level context influences what word (i.e., ordered set of phonemes) is heard by a listener. The basic technique is a simple extension of the added/replaced methodology to the sentence level, with one important modification. In the word-level experiments, the test words were selected in such a way that if a replaced phoneme were restored, a unique word (the original) would be formed. In the sentence level version, replacement items were always ambiguous; restoration could produce at least two different words. For example, the pair of words "battle" and "batter" formed a *basic word pair* in this experiment. In the replacement versions of these words the final (syllabic) liquids were replaced, leaving "bat*" in each case. The added or replaced versions of these words occurred in sentences such as "The soldier's thoughts of the dangerous bat* made him very nervous" or "The pitcher's thoughts of the dangerous bat* made him very nervous." In addition to the added/replaced judgment, subjects made a forced-choice *word judgment* (in this example, between "battle" and "batter"). The use of sentence pairs with multiply restorable words provides information on which phoneme is restored, in addition to the rate of restoration. Sherman (Note 1) first used multiply restorable stimuli and reported that the sentential context strongly influenced which phoneme was restored; this was true even if the critical sentential context occurred after the restoration item. Experiment 3 investigates restoration of multiply restorable words with the improved methodology used in the first two experiments.

Two theoretical issues are considered in Experiment 3. First, what effect does predictability of a word have on phonemic restoration? The results of the priming condition in Experiment 2 suggest that making a word predictable should increase restoration, causing a drop in discriminability of the added and replaced versions. However, the listeners' predictions in the present experiment are cued by semantic/syntactic information; it remains to be seen whether these sources of information affect restoration in the same way that lexical knowledge does. It is possible, for example, that these higher level information sources have effects at higher levels in the task. If so, predictability might have its primary effect on Beta, the index of postperceptual bias, rather than on the perceptual index d'.

To manipulate predictability of a word, quadruplets of sentences were used. For example, in addition to the two predictable sentences mentioned earlier for the "battle"-"batter" word pair, the unpredictable sentences "The soldier's thoughts of the dangerous batter made him very nervous" and

"The pitcher's thoughts of the dangerous battle made him very nervous" were included. In these sentences, the disambiguating cues ("pitcher's" and "soldier's") have been switched, making the critical word much less predictable.

The second factor considered in Experiment 3 is the role that memory load plays in the restoration illusion. Pisoni's (1973) work at the syllabic level has shown that speech is rapidly reencoded, with a loss of acoustic detail. A similar reencoding (perhaps at a higher level) could account for some of the observed restoration results. To test this possibility, Experiment 3 was run twice. In one version, the sentence stopped as soon as the critical word had been presented, and subjects made their two judgments immediately. Comparing this condition with one in which subjects delayed their responses until the sentence finished provides a measure of the role of memory factors in the restoration phenomenon.

In summary, Experiment 3 investigates two factors potentially affecting phonemic restoration in sentences: predictability of the critical word and memory load. In addition, the phone class and phoneme position factors were carried over from the preceding experiments, since they provided particularly interesting results.

Method

Stimuli. Twenty-seven *basic word pairs* were used to generate all of the sentences and fragments used in Experiment 3. The 27 pairs included three instances of each of the nine conditions created by factorially crossing three phone classes (liquid, stop, and vowel) and three positions within the word (initial, medial, and final). For example, there were 3 basic word pairs in which the added/replaced phoneme was an initial liquid—"liver-river," "loyal-royal," and "locket-rocket." All basic words were two syllables long.

For each of the 27 basic word pairs, 16 test stimuli were constructed. The 16 stimuli were broken down as follows: There were the two memory load *conditions* outlined above—the *full sentence* and *fragment* conditions. To satisfy the ambiguously restorable and predictability constraints, four *versions* of the sentences were used; each of the two basic words appeared with its own disambiguating context (e.g., battle-soldier) and with the context of its mate (e.g., battle-pitcher). For the battle-batter pair, the full sentences were as follows: the soldier's thoughts of the dangerous battle made him very nervous; the pitcher's thoughts of the dangerous batter made him very nervous; the pitcher's thoughts of the dangerous battle made him very nervous; and the soldier's thoughts of the dangerous batter made him very nervous. The fragments were as follows: the soldier's thoughts of the dangerous battle; the pitcher's thoughts of the dangerous batter; the pitcher's thoughts of the dangerous battle; and the soldier's thoughts of the dangerous batter. Each of these 8 stimuli was presented in both an added and a replaced form. Thus a total of 432 (27 × 16) stimulus items were used in Experiment 3.

For each word pair, the 16 test stimuli were constructed from four utterances recorded by the same speaker who produced the stimuli in the other experiments. The utterances were spoken at a normal conversational rate, with slightly reduced intonation. The fragment stimuli were constructed from the full-sentence items by editing out everything after the end of the word that contained the added/replaced segment. Added and replaced versions of the eight utterances were constructed using the same technique and equipment used to construct the stimuli in the first two experiments. Table 7 presents the utterance lengths, critical segment onset times, segment lengths, and segment DRMSA for the fragment and full-sentence conditions.

Apparatus and procedure. On each trial, the word READY was displayed on a screen in front of the subject for 1 sec. This warning signal disappeared, and 300 msec later a test utterance was played binaurally over stereo headphones. When the utterance finished, the words ADDED and REPLACED immediately appeared on the left and right sides of the screen, respectively. Subjects made the usual added/replaced judgment as quickly and as accurately as possible, using the display of ADDED and REPLACED as a signal to respond. Seven hundred msec after all (1-4) of the subjects had responded, two words appeared on the screen, replacing ADDED and REPLACED. The two words were the basic pair appropriate for the utterance the subjects had just heard (e.g., "battle-batter"). Subjects were told to press one of two keys to indicate which of the two words they thought had been in the utterance on that trial (left key - left word, right key - right word). For both the added/replaced judgment and the word judgment, subjects were told to guess if necessary—both responses were required on all trials. One second after all of the subjects had made the word judgment, the next trial began.

The apparatus was the same as that used in the previous experiments with the following exceptions: (a) A Tektronix 602 display scope was used, and (b) subjects used response keys instead of response buttons. The same two keys were used for the added/replaced and word judgments (left - ADDED for first judgment, and left - left word for the second).

Each subject received 94 utterances, divided into two passes through 47 utterances. The first 20 items in each pass were practice and were not scored. The remaining 27 utterances included 1 utterance from each of the basic word pair sets (see the Design section below). The order of the 27 test items was determined randomly. On the first pass through the 47 utterances, either the added or replaced version of each utterance was randomly ($p = .5$) selected. The second pass through the items included the versions not presented on the first pass. The comparison of added and replaced versions was thus within subject. On each trial, the position (left or right) of the forced-choice words was randomly determined.

Table 7
Stimulus Information in Experiment 3

Stimulus type	Utterance length	Segment onset time	Segment length	Segment amplitude
Full sentence	2659	1802	137	671
Fragment	2063	1802	137	671
Phone Class				
Liquid	2957	1960	139	749
Stop	2566	1742	133	265
Vowel	2453	1702	138	1000
Position				
Initial	2624	1519	142	666
Medial	2626	1845	114	883
Final	2726	2041	154	466

Note. The utterance length, segment onset time, and segment lengths are in milliseconds. The segment amplitude is the digital RMS amplitude (DRMSA). Unless otherwise noted, the utterance lengths are for the full sentence condition. The DRMSA was normalized such that vowels had a mean of 1000.

The data for each subject were collected in a single session lasting approximately 20 min.

Design. The fragment and full-sentence conditions were each run individually—no subject received both kinds of utterances. Each subject heard one added/replaced item from each of the 27 basic word pairs. Since there are four utterances per basic word pair set in each of the conditions, four subjects were needed to span the entire stimulus set. Thus four test versions of each memory load condition were constructed. Within each test version a Latin square procedure was used to counterbalance the four versions of each condition, the phone class, and the phone position of the added/replaced phone.

Subjects. A total of 80 subjects participated in Experiment 3, 40 in each memory load condition. The subjects were randomly assigned to the four versions of each test, yielding 10 subjects in each test version.

All subjects were native English speakers with no known hearing problems. They received $2 or course credit for their participation.

Results

Measures. For the added/replaced judgments, the two primary measures used in the first two experiments, d' and Beta, were used again. For the forced-choice word judgments, a similar signal detection theory breakdown can be used. The primary measure of interest is Beta, which is an index of how much context biases the listener to hear (or at least report) the word that fits. Beta measures the likelihood that the predicted word was reported, regardless of what word was originally there.

Composite subjects. Recall that in each memory load condition, 4 subjects were required to span the stimulus set. In each of these conditions, formation of 10 *composite subjects* (pseudosubjects who received all the stimuli in a condition) was done as follows: for each subject, the Beta score for the added/replaced judgment was computed. The 10 subjects who had received a given version of the experiment were ordered according to their Beta scores. Composite Subject 1 was then formed by combining the data from the subject within each group of 10 who had the lowest Betas. Composite Subject 2 was comprised of those with the second lowest Betas, and so on, up to Composite Subject 10, those with the highest Betas. Combining the data in this way has the virtue of putting together data from subjects with similar decision criteria on the added/replaced judgment, decreasing within-pseudosubject variance.[5]

The major issues being considered in Experiment 3 are the roles that semantic–syntactic context and memory load play in phonemic restoration. Table 8 presents the added/replaced judgment data, which reflect how these factors affect the rate of restoration. Two aspects of the d' data are most important. First, there is no difference in performance as a function of when the re-

[5] In addition to providing pseudosubjects who received all of the stimuli, this procedure increases the number of observations per cell for each pseudosubject. As noted before, the d' and Beta measures are highly nonlinear for extreme miss and false alarm rates; the larger cell sizes produced by this procedure substantially reduce such extreme values and thus increase statistical power.

Table 8
Memory Load by Predictability Breakdown for Experiment 3

Context condition	Fragment		Full sentence	
	d'	Beta	*d'*	Beta
Predicted	1.48	3.07	1.71	2.02
Unpredicted	1.42	1.33	1.34	1.33

Note. The *d'* and Beta values are in absolute units.

Table 9
Bias (Beta) on Word Judgments as a Function of Memory Load and Restoration Occurrence

Restoration occurrence	Fragment	Full sentence
Restoration	.67	.77
No restoration	.83	.86

Note. The Beta values are in absolute units.

sponse is made; memory load appears to play no role. Second, and more surprising, discriminability is actually better in predictable situations than in unpredictable ones. A two-way analysis of variance (Memory Load × Predictability) bears out these observations; memory load, $F(1, 18) = .10, p > .20$; predictability, $F(1, 18) = 4.38, p = .06$. Although the predictability effect is larger for full sentences than for fragments, the interaction is not reliable, $F(1, 18) = 2.30, p > .10$. Note that the predictability result is in direct opposition to the predictions of the expectation-driven component of the interactive schema model; it also conflicts with the priming results. Apparently, high-level information of the type available in this testing situation is unable to influence the low-level perceptual process.

The Beta values in Table 8 show that semantic–syntactic context is by no means impotent, however. There is a large bias toward reporting predictable words as intact. In a two-way analysis of variance of the Beta scores (Memory Load × Predictability), the predictability effect was marginally significant, $F(1, 18) = 3.98, .05 < p < .10$. However, the effect was much more reliable than this analysis suggests—19 of the 20 pseudosubjects showed it. Thus, although discriminability was unimpaired by predictability, listeners did report predictable words to be intact more than unpredictable ones; less "psychological intactness" is required when a word is expected. As in the *d'* data, memory load had no effect, $F(1, 18) = .43, p > .20$.

Table 9 presents data that bear on the question of whether semantic–syntactic context determines *what* is restored. The values are Beta scores for the forced-choice word judgments. All the data in the table come from trials on which the stimulus was a replacement item. The scores in the top row come from trials on which subjects responded "added" to the replacement item (i.e., restored either perceptually or postperceptually); the bottom row data come from the trials on which subjects correctly reported the replaced stimulus as replaced. The Beta scores were set up such that a value less than 1.0 indicated a bias toward reporting the word predicted by context. The first thing to notice is that in all conditions, there is a bias toward reporting the predicted word. However, this bias is not uniform; on trials in which restoration occurred, listeners were more biased toward the predicted word than when no restoration occurred. An analysis of variance (Restoration × Memory Load) revealed that this difference was marginally significant, $F(1, 18) = 3.90, .05 < p < .10$. It also indicated that the tendency toward more bias in the immediate response condition was not significant, $F(1, 18) = 1.63, p > .10$.

It is of course possible to compute a *d'* score for the word judgments as well as the Betas that have been discussed. The *d'* measure reflects the subjects' ability to discriminate between replacement items that were constructed from different original words (e.g., "bat*" from "battle" and "bat*" from "batter"). These values were generally quite low, averaging about .7, indicating that few (but some) cues to the original remained.

Table 10 presents the *d'* and Beta scores (from the added/replaced judgment) for the fragment and full-sentence conditions, broken down by phone class of the added/replaced phone. The values reported are the means for the 10 composite subjects in each condition. Two-way analyses of variance (Phone Class × Memory Load Condition) were conducted on the *d'* and Beta scores in

Table 10
Added/Replaced Judgments: Phone Class Breakdown

	Liquid	Stop	Vowel
Fragment			
d'	1.53	1.29	1.61
Beta	2.78	3.39	1.65
Full sentence			
d'	1.64	1.23	1.79
Beta	1.53	1.86	1.46

Note. The *d'* and Beta values are in absolute units.

Table 11
Added/Replaced Judgments: Phone Position Breakdown

	Initial	Medial	Final
Fragment			
d'	1.46	1.19	1.68
Beta	2.35	1.37	2.00
Full sentence			
d'	1.77	1.39	1.54
Beta	2.53	1.15	2.99

Note. The *d'* and Beta values are in absolute units.

Table 10. The effect of phone class was significant on the measure of replacement detectability, $F(2, 36) = 6.76, p < .01$, but not on Beta, $F(2, 36) = 1.70, p > .10$. The effect of memory load was not significant on either measure (for both, $F \leq 1$). The potency of the phone class factor indicates that the strong influence of bottom-up confirmation found in the first two experiments (using isolated words) is not lost in a situation with more top-down influences. As in isolated words, the phone class manipulation has its impact at the perceptual level, not on higher level decision processes. The same ordering of restorability is found in both situations—for both fragments and full sentences, stops are best restored, followed by liquids and vowels.

The Beta scores for words in sentential context are higher than for words in isolation (see Table 2), indicating that subjects are generally more likely to report the words as intact in context. However, the *d'* measure indicates better detection of phoneme replacement in sentences (or fragments) than in isolation. A likely basis for this surprising result is the influence of prosody (especially pitch contour) in sentences. The replacement version of a sentence will often introduce a break into the prosodic structure, whereas the added version seldom does. This provides listeners with an artifactual cue for discriminating added and replaced versions in sentences. Thus we cannot compare absolute scores across Experiment 3 and Experiments 1 and 2. However, within each experiment, we may safely consider the patterns of results.

Table 11 presents the *d'* and Beta scores for the fragment and full-sentence conditions, broken down by phoneme position within the target word. A comparison of these data with the comparable results using isolated words (Table 2) reveals that the patterns differ somewhat. In isolation, no difference in discriminability was found as a function of position. In sentences, although initial and final position are roughly equal, medial position is consistently poorer. A two-way analysis of variance (Phone Position × Memory Load) indicates that this position effect in sentences is reliable, $F(2, 36) = 4.79$, $p < .05$. The bias measure shows an even more striking change. In isolation, subjects were most biased toward reporting replaced in initial position and most biased toward added in medial position (see Table 2). In sentences, medial segments were most likely to be called replaced, Beta, $F(2, 36) = 3.60, p < .05$. This result may be due to the prosody cue discussed earlier. The introduction of a noise burst in the middle of a word (whether added to or replacing a phone) is likely to create the impression of a break in structure; this impression is less likely with noise bursts at word boundaries (i.e., word initial or word final), since breaks between words are expected (or even created) by the listener. If subjects are using the prosody-break cue, this would lead to the lower medial Betas directly and to lower medial *d'*s by use of a cue that is unreliable in this situation. As in the previous analyses, memory load had no effect on the added/replaced judgments (for both, $F < 1$).

Discussion

The central question under consideration in Experiment 3 was what role sentential context plays in phonemic restoration. The answer appears to be that context serves to bias listeners. For the word judgments, this result is not at all surprising; Sherman (Note 1) has shown that subjects' reports of words tend to follow the sentential context. However, the bias-based nature of context effects on the added/replaced judgment is much more surprising and much more important. The data indicate that listeners are more inclined to report words as intact if the words are predictable from preceding context than if they are not. This bias effect was coupled with an *improvement* in discriminability for predictable words.

These data place an important constraint on the interactive schema model of speech perception that was posited at the outset. The failure to find a discriminability decrement for predictable words suggests that the higher level (syntactic-semantic) information was not being passed down to the lower phonetic-phonological level; the top-down processing that was evident between the lexical and phonetic-phonological level did not occur here. This is not to say that the information was not used. On the contrary, the Beta data show clearly that high-level information is effective. However, its effect is at higher levels, particularly the decision-making level. The data also do not rule out the possibility of top-down use of sentential context. It is possible that in an extremely predictive context, a discriminability decrement might be found. However, with the materials tested (and with most normal discourse), the level of prediction was insufficient to activate a particular lexical entry strongly enough to produce the performance decrement found in the priming situation of Experiment 2.

In fact, discriminability actually improved for predictable words. This improvement suggests that the various levels of processing share some common resources. When preceding context makes a word predictable, the load on the perceptual system decreases. This apparently leaves more processing capacity available for the fine level of acoustic analysis needed to discriminate added and replacement items.

In light of this finding, the results of the memory load manipulation are a little bit surprising. On essentially all measures, requiring subjects to delay their responses had no effect. Maintaining a response in memory should place some load on the system, hurting performance. Either this additional load is negligible, or it is balanced by some aspect of the delay condition, such as the greater naturalness of whole sentences.

An overall comparison of performance on isolated words (Experiment 1) and words in sentences (Experiment 3) revealed higher discriminability in sentences. The pattern of results for critical phonemes in different positions within words suggested that subjects were using pitch contour breaks as a cue to improve discriminability. The possible sensitivity of the paradigm to prosodic factors is both a plus and a minus; although it complicates the study of segmental effects, it may provide insights into suprasegmental processing.

There is a second factor that may have contributed to the relatively good discriminability in Experiment 3. Recall that in the first two experiments, the words were chosen so that a unique lexical item could be restored. In Experiment 3, the critical words were designed to be multiply restorable in order to investigate context effects. It is possible that the existence of more than one potential percept inhibited restoration, improving discriminability. If the purpose of the restoration mechanism is to reconstruct the original word spoken, the system might be stymied when more than one acceptable candidate is available. Research is currently under way to test this hypothesis.

Finally, the results of the phone class manipulation were straightforward—stops were most restorable, followed by liquids and vowels. This pattern of results is identical to that observed for isolated words. It thus appears that although the functioning of top-down expectations differs in the two situations, the process of bottom-up confirmation of hypotheses is similar.

General Discussion

The three experiments were designed to provide some insight into the encoding of the sounds of speech into meaningful units. The focus of this inquiry was to determine how much of this encoding is a function of the incoming sounds and how much of it is determined by the knowledge of the listener. The results, not surprisingly, indicate that speech perception depends on the interaction of both the top-down expectations generated by the listener's knowledge and the bottom-up confirmation provided by the acoustics of the signal. These bottom-up influences were actually more potent than one might have expected, given the fundamentally knowledge-driven nature of a phenomenon like phonemic restoration; a very potent factor in whether restoration occurred turned out to be the phone class of the speech segment to be restored and its acoustic similarity to the replacement sound. Samuel (1981) has shown that specific bottom-up characteristics of the signal, including its amplitude, continuity, and periodicity, have significant effects on the rate of restoration.

The technique of separating restoration into discriminability and bias revealed that various higher level knowledge sources may increase the miss rate in different ways. Lexical knowledge can be brought to bear during the perceptual process in a truly top-down fashion. Thus more restoration occurred in words than in phonologically legal pseudowords; the difference resulted in poorer *d*'s for words than pseudowords. Similarly, increasing the availability of lexical information, through priming, reduced discriminability of added and replaced stimuli. In contrast, sentential information produced higher miss rates through a bias effect; listeners were more willing to say "intact" when a word was expected than when it was not. Discriminability actually improved for sententially predictable words, suggesting that the load on the perceptual system was reduced through predictability.

The model of speech perception that emerges from these results is a modified version of Rumelhart's (1977) interactive schema model. The critical feature of Rumelhart's model is the combination of both bottom-up and top-down modes of processing. The results of Experiment 2 demonstrate the powerful role of top-down processing in the perception of speech. The results of Experiment 3 place an important constraint on the interactive schema model—not all sources of information appear to pass their results in both directions. In particular, the data indicate that high-level syntactic-semantic information is used in the high-level decision process, but it does not reach the acoustic-phonetic level. This result suggests that in studieş in which the perceptual and decision stage components are inseparable (e.g., Foss & Blank, 1980), the observed contextual effects are due to decision stage factors.

This model of speech perception may be assessed by considering how well it agrees with the results of several recent studies in which different methodological approaches were taken. Each of these approaches solves some of the problems inherent in the methodology first used to study phonemic restoration.

Cole (1973; Cole & Jakimik, 1979) introduced a technique in which subjects listen to a passage that contains occasional mispronunciations. The task was to press a button whenever a mispronunciation was detected. Cole varied the degree of mispronunciation (one, two, or four distinctive features) and its position within the word (word initial, second syllable final, or word final). The mispronunciation miss rate, an index of restoration, varied with similarity of the expected and actual phonemes but was unaffected by position of the mispronunciation. The similarity of the mispronunciation to the original is a bottom-up factor that appears to be quite potent (as in the present study). Reaction times for the detections also varied with similarity. In addition, reaction times varied with position; responses were relatively slow for word-initial mispronunciations, whereas second-syllable and word-final responses were about equally fast. Cole suggested that the long reaction times for word-initial mispronunciations reflect the dominance of initial phonemes in lexical access—a mispronunciation there leads to erroneous entry into the lexicon.

The results of the position factor in Experiments 1 and 2 generally confirm Cole's hypothesis. For words in isolation, there was a consistent tendency for subjects to report something missing if an extraneous sound occurred at the beginning of a word, regardless of whether the sound coincided with or replaced the initial phone. Such a diffential sensitivity tends to support Cole's view. The priming data provide support for a refinement of the special role of word beginnings. Those data showed that when expectations exist, the initial part of a word can *confirm* the expectation, producing greater restoration of later phones. The failure to obtain the effect in Experiment 3 may well be due to the artifactual prosody cue. Thus the results of the present study may be taken as support for a special confirmatory role of word-initial phonemes.

Marslen-Wilson (1975) introduced another variant of restoration methodology, the shadowing technique. Marslen-Wilson had subjects shadow (repeat almost simultaneously) speech in which mispronunciations occurred, and as in Cole's procedure, inferred restoration when a mispronunciation was not repeated (i.e., the subject restored the proper pronunciation). Marslen-Wilson found that most restoration occurred in the last two syllables of words that the subjects expected. This result suggests that restoration was initiated by the sentential context (i.e., expectations generated by listener knowledge) and was finalized when the initial syllable confirmed the contextually-driven hypothesis.

Marslen-Wilson and Welsh (1978) compared Cole's (1973) mispronunciation technique and the shadowing method, using the same recorded material for each. The results for the two procedures were similar. The biggest difference was that for grossly mispronounced phonemes (three distinctive features changed), subjects rarely restored using the mispronunciation task (6%) but were fairly likely to restore when shadowing (24%). This result may be due to subjects focusing their attention on mispronunciations in Cole's task, or to the great attentional demands of shadowing. In any event, degree of mispronunciation was a very potent factor on both measures. As in Marslen-Wilson's (1975) study, predictability of a word greatly increased the likelihood of a shadowing restoration. In the present study, sentential predictability also increased "restoration," but did so by biasing the decision process. In the shadowing paradigm, there is no way to discover whether the higher restoration rate was due to a perceptual effect or some sort of postperceptual decision stage.

Marslen-Wilson and Welsh's (1978) model of lexical access from speech incorporates both bottom-up and top-down processing. They argued that the incoming acoustic information suggests a group of word candidates (the *cohort*), and that higher level factors help to choose from among the candidates. Note that this model reverses the roles of bottom-up and top-down information; the bottom-up information creates the expectation (or cohort), and the higher level knowledge is used to confirm one of the candidates. It is not clear how to choose between the two formulations.

Bashford and Warren (1979) have recently introduced a restoration methodology that seems particularly promising for investigating high-level effects on restoration. In this procedure, a speech passage is interrupted in such a way that the listener only hears half of the passage. For example, the first 500 msec of the passage might be heard, the next 500 msec deleted, the next 500 msec heard, and so forth. Miller and Licklider (1950), using word lists, had reported that replacing the deleted portions with broadband noise produced speech that sounded more complete than the speech produced by leaving the deleted segments as silence. Since Miller and Licklider found no corresponding increase in intelligibility, the illusory continuity appears to be similar to the bias component of restoration observed in Experiment 3. The continuity effect also has a property that would be expected if postperceptual restoration were its basis—the continuity breaks down if the interruptions are too long, reducing the necessary context.

From these facts, Bashford and Warren (1979) developed a new measure of restoration: the point at which the continuity effect appears (or disappears) when the rate of speech-noise switching is varied. Subjects were given control over this rate, and made threshold judgments for the continuity effect. Bashford and Warren found that if broadband noise was the replacement sound, listeners heard continuous speech with noise-filled deletions as long as 304 msec. When silence was left, the continuity broke down at 52 msec; silence did not confirm the expectations generated by the remaining speech. The results of these and other conditions in Bashford and Warren's study led them to conclude that the success of restoration depends on (a) the spectral similarity of the replacement sound and the speech it replaces, and (b) the amount of linguistic context available. In short, restoration is a function of expectation and confirmation.[6]

The results of three quite different methodologies thus coverge with those obtained in Experiments 1-3. The strengths and weaknesses of the added/replaced paradigm nicely complement those of the alternative methods. The strongest virtue of the procedure used in the present study is its ability to separate perceptual discriminability from bias; none of the other three methods can do this because none has any usable false alarm rate. Thus bias and discriminability (of a restored and a "real" phoneme) are inextricably intertwined.

A second virtue of the procedure is that it does not impose a second, high-resource task on the listener, as is the case in the shadowing paradigm. The shadowing task was first used precisely because it did absorb most of the subject's attentional capacity. With this strain on the system, it is likely that the listener is forced to rely more on top-down processing (knowledge already in hand) than in the normal situation. Thus the shadowing methodology probably exaggerates the role of expectations. In contrast, Cole's (1973) mispronunciation task and the added/replaced task focus the listener's attention on the speech sounds, accentuating bottom-up processing to some degree. The middle ground between the models derived from the shadowing and detection paradigms is therefore probably closest to the true functioning of the speech system.

[6] Restoration of speech sounds is not necessarily unique in being a function of expectation and confirmation—the interactive schema model of Rumelhart (1977) is applicable to perception in general. In fact, Warren, Obusek, and Ackroff (1972) have reported a nonspeech analogue to phonemic restoration that appears to follow similar rules, including sensitivity to such confirmatory factors as spectrum and intensity.

Reference Note

1. Sherman, G. *The phonemic restoration effect: An insight into the mechanisms of speech perception.* Unpublished master's thesis, University of Wisconsin—Milwaukee, 1971.

References

Bashford, J., & Warren, R. M. Perceptual synthesis of deleted phonemes. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers.* New York: Acoustical Society of America, 1979.

Cole, R. Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics,* 1973, *11,* 153-156.

Cole, R., & Jakimik, J. A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech.* Hillsdale, N.J.: Erlbaum, 1979.

Foss, D., & Blank, M. Identifying the speech codes. *Cognitive Psychology,* 1980, *12,* 1-31.

Kucera, H., & Francis, W. *Computational analysis of present-day American English.* Providence, R.I.: Brown University Press, 1967.

Layton, B. Differential effects of two nonspeech sounds on phonemic restoration. *Bulletin of the Psychonomic Society,* 1975, *6,* 487-490.

Marslen-Wilson, W. Sentence perception as an interactive parallel process. *Science,* 1975, *189,* 226-228.

Marslen-Wilson, W., & Tyler, L. The temporal structure of spoken language understanding. *Cognition,* 1980, *8,* 1-71.

Marslen-Wilson, W., & Welsh, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology,* 1978, *10,* 29-63.

Miller, G. A. & Licklider, J. C. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America,* 1950, *22,* 167-173.

Obusek, C., & Warren, R. M. Relation of the verbal transformation and the phonemic restoration effects. *Cognitive Psychology,* 1973, *5,* 97-107.

Pisoni, D. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics,* 1973, *13,* 253-260.

Rumelhart, D. E. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI.* Hillsdale, N.J.: Erlbaum, 1977.

Samuel, A. G. The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance,* 1981, *5,* 1124-1131.

Taft, M. Lexical access via an orthographic code: The basic orthographic syllabic structure (BOSS). *Journal of Verbal Learning and Verbal Behavior,* 1979, *18,* 21-40.

Warren, R. M. Illusory changes of distinct speech upon repetition—The verbal transformation effect. *British Journal of Psychology,* 1961, *52,* 249-258.

Warren, R. M. Perceptual restoration of missing speech sounds. *Science,* 1970, *167,* 392-393.

Warren, R. M., & Obusek, C. Speech perception and phonemic restorations. *Perception & Psychophysics,* 1971, *9,* 358-363.

Warren, R. M., Obusek, C., & Ackroff, J. Auditory induction: Perceptual synthesis of absent sounds. *Science,* 1972, *176,* 1149-1151.

Warren, R. M., & Sherman, G. Phonemic restorations based on subsequent context. *Perception & Psychophysics,* 1974, *16,* 150-156.

Received August 12, 1980 ■