789

Perceptual restoration of a "missing" speech sound: Auditory induction or illusion?

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut

This study investigated whether the apparent completeness of the acoustic speech signal during phonemic restoration derives from a process of auditory induction (Warren, 1984) or segregation, or whether it is an auditory illusion that accompanies the completion of an abstract phonological representation. Specifically, five experiments tested the prediction of the auditory induction (segregation) hypothesis that active perceptual restoration of an [s] noise that has been replaced with an extraneous noise would use up a portion of that noise's high-frequency energy and consequently change the perceived pitch (timbre, brightness) of the extraneous noise. Listeners were required to compare the pitch of a target noise, which replaced a fricative noise in a sentence, with that of a probe noise preceding or following the speech. In the first two experiments, a significant tendency was found in favor of the auditory induction hypothesis, although the effect was small and may have been caused by variations in acoustic context. In the following three experiments, a larger variety of stimuli were used and context was controlled more carefully; this yielded negative results. Phoneme identification responses collected in the same experiments, as well as informal observations about the quality of the restored phoneme, suggested that restoration of a fricative phone distinct from the extraneous noise did not occur; rather, the spectrum of the extraneous noise itself influenced phoneme identification. These results suggest that the apparent auditory restoration which accompanies phonemic restoration is illusory, and that the schema-guided process of phoneme restoration does not interact with auditory processing.

One of the best known and most interesting findings in speech perception research is the "phonemic restoration illusion" first demonstrated by Warren (1970; Warren & Obusek, 1971; Warren & Sherman, 1974). In Warren's original experiments, subjects listened to a sentence in which the acoustic signal pertaining to one phoneme, the fricative /s/, had been excised and replaced with an extraneous sound or with silence. When an extraneous sound, such as a cough, was present, listeners claimed to hear the speech as intact and could not localize the cough (or the speech sound masked by it) accurately within the sentence. The silence, however, was usually heard as replacing the /s/. Warren (1976, 1984) interpreted these findings as reflecting general principles of auditory perception: "Phonemic restoration can be considered as a verbal form of heterophonic auditory induction, and if the neural units stimulated by the extraneous sound include those which would be stimulated by the missing phoneme (as identified through contextual cues), then perceptual synthesis of this phoneme may occur" (1976, p. 403). This perceptual synthesis, however, is "governed by linguistic rules" (1984, p. 375).

Later research by Samuel (1981a, 1981b) confirmed that the extraneous sound must have some spectral resemblance to the missing speech sound for phonemic restoration to be effective. Samuel's innovative contribution was methodological: Instead of relying merely on subjects' inability to identify the "missing" sound, he employed a one-interval discrimination paradigm in which subjects had to distinguish between speech stimuli in which the extraneous sound *replaced* part of the speech signal, and stimuli in which the extraneous sound was merely *added* to the critical speech segment. The relative difficulty of this discrimination was considered a measure of the strength of phonemic restoration, the reasoning being that "replaced" stimuli cause auditory synthesis of the missing phoneme and therefore should sound just like "added" stimuli if restoration is complete. The discriminability and bias indices of signal detection theory (d' and β) were the dependent variables in a number of subsequent investigations of various conditions that might influence the strength of the phonemic restoration effect (Samuel, 1981a, 1981b, 1987, 1991; Samuel & Ressler, 1986; Trout & Poser, 1990).

Two Hypotheses

If a "replaced" stretch of the speech signal appears to be present in the listener's subjective auditory experience, where does the auditory information for that restoration come from? One possibility is that the restoration is an auditory illusion that arises from context-induced phonological completion and consequent phonetic imagery (i.e.,

14

inner speech). That is, perhaps subjects do not (or cannot) actually make judgments about the intactness of the auditory input signal, and instead report on an abstract phonological or imaginary phonetic representation that is activated in the process of word recognition (cf. Foss & Blank, 1980). In that case, the extraneous sound replacing the missing phoneme provides "bottom-up confirmation" (Samuel, 1981b), presumably by facilitating lexical access to the extent that it masks the absence of the missing phoneme, but otherwise it does not participate in the restoration process. The strength of the illusion should then depend on the success and speed of lexical access, and this hypothesis is indeed supported by Samuel's findings that the illusion is reduced for nonwords (Samuel, 1981a) and lexically ambiguous words (Samuel, 1987), whereas it is enhanced by preexposure to nonwords (Samuel, 1981a). It is further supported by the result (Samuel, 1991; Samuel & Ressler, 1986) that cuing subjects' attention to the critical phoneme in the word reduces the illusion, which suggests that the speech input is not perceived as complete when its auditory memory trace is examined closely. However, as is well known from categorical perception research (see Repp, 1984), access to detailed auditory properties of speech is quite difficult in high-uncertainty listening situations, so the absence of genuine auditory restoration may go unnoticed when the phonological structure is complete.

Another possibility, however, becomes evident when the "added" condition is considered, in which the original acoustic information supporting perception of the critical phoneme is actually present in the input signal. In this condition, it seems plausible that the listener's auditory system would segregate the extraneous sound from the speech, so that the speech signal is heard as intact because it *is* intact. Since the auditory system does not "know" in advance whether the original speech information is present or absent, one might suppose that it would attempt the same context-guided segregation procedure in the "replaced" condition. Since the extraneous sound must be spectrally similar to (i.e., a potential masker of) the missing speech sound for restoration to occur (Samuel, 1981a, 1981b; Warren, 1984), it may be argued that the replaced speech sound is not really missing; rather, it is hidden in the extraneous sound and must be extracted and separated from it. This separation may be less effective or complete than in the case of "added" stimuli—only some resemblance of the missing speech sound may be restored—but the basic process may be the same. According to this reasoning, top-down expectations guide the listeners as to what to extract, but these expectations interact with primitive auditory processes that attempt to reconstitute the speech signal. Phonemic restoration is not an illusion in that case; it is an instance of auditory source segregation, and the acoustic signal does not simply confirm a restoration accomplished at a higher level but provides the material out of which the restored

sound is carved by the auditory system. This is essentially Warren's (1984) "auditory induction" account.

These two possibilities have implications not only for the perception of the restored speech segment, but also for the perception of the extraneous sound. According to the first hypothesis (henceforth, the *top-down completion* hypothesis), the restored speech should "sound" quite normal because subjects report on an ideal internal image accompanying the restored phoneme, and the extraneous sound should sound very similar to the way it sounds when it occurs in isolation. According to the second hypothesis (the *segregation* hypothesis), the extraneous sound is partitioned into a likeness of the acoustic speech segment expected at that point and into a residue that is perceived as an extraneous sound. Because the acoustic energy necessary to support perception of the speech segment is subtracted from the extraneous sound, as it were, the perceived sound quality (as well as the loudness) of the extraneous sound should change when restoration occurs. Likewise, the sound of the restored speech segment may not be fully natural if the extraneous sound provides only a partial "spectral blanket." Because listeners find it generally very difficult to judge the auditory quality of speech segments, particularly in a sentence context, the present study focused on the predictions concerning the perception of the extraneous sound.

In addition to Warren's and Samuel's work, two other contemporary theories are directly relevant to the hypotheses contrasted here.[1] In his thoughtful and comprehensive discussion of auditory scene analysis, Bregman (1990) distinguishes between primitive and schema-driven processes. Phonemic restoration is clearly schema driven: The acoustic–phonetic context alone is hardly sufficient to suggest the missing speech segment to a listener. Linguistic knowledge, or, in Bregman's terminology, a set of speech schemata, is needed to infer the identity of the missing segment from the context. The question posed here is whether this schema-driven process interacts with primitive auditory processes of source assignment and timbre perception. Bregman believes that it does not: "Speech schemas are the same as other recognition schemas in the following way: when they make use of the information that they need from a mixture, they do not remove it from the array of information that other description-building processes can use" (p. 638). And, with specific reference to phonemic restoration, he states that it is "based on schema-governed stream segregation and for this reason will not give rise to a distinct residual" (p. 372). According to his theory, a residue-forming process of auditory stream segregation would occur only when the segregation is directly induced by the surrounding acoustic context, as demonstrated in spectrally complex but steady-state nonspeech sounds whose spectral components may be "captured" by a preceding and following simpler sound according to the law of continuity (e.g., Bregman, 1990; Bregman & Pinker, 1978; Warren,

Obusek, & Ackroff, 1972). Thus, Bregman's view, like Samuel's, seems to be closer to the top-down completion hypothesis than to the segregation hypothesis.

In contrast, Mattingly and Liberman (1988, 1990; also Liberman & Mattingly, 1989) have proposed that the speech perception system "preempts" auditory scene analysis, taking what it needs from the auditory signal and leaving a residue for general auditory processes to deal with. This hypothesis is based primarily on evidence from the duplex perception paradigm, which has been investigated intensively by Liberman and his colleagues (see, e.g., Mann & Liberman, 1983; Whalen & Liberman, 1987). There are some interesting parallels between duplex perception and phonemic restoration. The dichotic duplex perception paradigm juxtaposes an ambiguous syllable lacking a critical formant transition (the "base") in one ear with the isolated formant transition in the other ear. Listeners report hearing the intact syllable as well as the chirp-like isolated transition, just as they claim to hear both the intact speech and the extraneous sound in the phonemic restoration paradigm. In duplex perception, as in phonemic restoration, the question then arises whether the full syllable is constructed at a higher level (by combining phonetic elements perceived as coming from separate sources, in this instance), so that its perceived auditory presence in lieu of the base is illusory (cf. Nusbaum, Schwab, & Sawusch, 1983), or whether the missing transition is actively segregated from the spectral information entering the opposite ear, at an early stage in binaural processing.

Whalen and Liberman (1987) clearly favor the latter interpretation and have provided some empirical support for it in a monaural duplex perception situation. By varying the relative amplitude of a sinusoidal frequency transition embedded in a syllable, they determined that the detection threshold for the transition as a whistle was much higher than the phoneme identification threshold. That is, the transition contributed to speech perception even when it could not be heard as a separate auditory event. A similar finding for the standard dichotic duplex perception paradigm was reported by Bentin and Mann (1983, 1990). On the basis of these results, Whalen and Liberman (1987) and Mattingly and Liberman (1988) suggested that the speech system takes the auditory information it needs and leaves a less intense frequency transition for the general auditory system, too weak to be detected unless its intensity is raised substantially. Applied to the phonemic restoration paradigm, this view might predict that the speech system will grab the information it needs from the acoustic spectrum provided by the extraneous sound, leaving an impoverished residue. This speech-specific "preemptiveness hypothesis" thus coincides with the (more generally schema-based) segregation hypothesis, proposed above.[2]

Methodological Considerations

The present study is about *phone restoration*, more than about phoneme restoration. The *phoneme*, an abstract phonological category without a definite sound, is distin-

guished here from the *phone*, a concrete auditory percept based on a relatively homogeneous acoustic speech segment and having elementary auditory properties such as pitch, timbre, and duration (see Repp, 1981a). Whereas in the phoneme restoration paradigm all acoustic cues for a given phoneme are excised from the acoustic signal, in the present study only one acoustic segment was removed—specifically, the noise portion of a fricative consonant. The approach taken here required that the critical segment be acoustically as simple as possible, so that the effects of its segregation from an equally simple extraneous sound could be predicted, perceived, and measured. Other cues to fricative identity—namely, the vocalic formant transitions preceding and following the noise portion—remained intact and, without any doubt, enhanced restoration of the correct fricative phoneme (cf. Mann & Repp, 1980; Mann & Soli, 1991). Since fricative phoneme restoration has been demonstrated numerous times in speech signals harboring no residual fricative cues, the occurrence of phoneme restoration in the presence of such cues was taken for granted.

Given that there are residual fricative cues in the surrounding speech signal, auditory restoration of a fricative noise (if it occurs) may be guided not only "from above" by a phonetic speech schema but also "from below" by local acoustic properties that call for a "good continuation" through the gap covered up by the extraneous noise. Soli (1981), for example, has demonstrated that formant transitions extend into and out of fricative noises and thus provide acoustic continuity with the surrounding vocalic segments. There may even be low-amplitude noise or breathiness in the immediately adjacent vocalic segments, due to the change in laryngeal action between voiced and voiceless segments, which may induce restoration of a noisy speech segment. Although the acoustic context alone cannot tell the auditory system the precise spectrum and amplitude of the missing phone (information that a speech schema might provide), it does provide pointers to some acoustic properties of that noise. Thus, conditions in the present experiments were maximally conducive to auditory restoration and segregation of the "missing" phone: Given the guidance provided by the acoustic context, both at the level of primitive auditory scene analysis and indirectly via the top-down influence of lexical/phonological units stimulated by the residual cues, the restored fricative noise should be partialed out from whatever sound occurs in its place, if this is at all what happens in phoneme restoration. Only if phoneme restoration is a pure top-down process (i.e., never "consumes" acoustic energy) would the experimental results be expected to favor the top-down completion hypothesis.

The experimental task required the subjects to pay attention to the extraneous sound (a noise) that replaced the target phone (typically [s]) in a brief sentence and to compare its auditory quality to that of a probe noise occurring either before or after the sentence. According to the top-down completion hypothesis, the target noise should be perceived veridically: If the target and probe noises

are physically identical, they should sound the same. According to the segregation hypothesis, however, physically identical target and probe noises should not be perceived as identical. The nature and direction of the perceived difference should reflect the acoustic properties of the restored speech sound: Since an [s] noise is characterized by strong high-frequency energy, the residue of the target noise should be depleted of energy in that region. Noises can be described as having a pitch-like quality or relative "brightness" of timbre that depends on their spectral composition (see, e.g., Glave, 1973; Hesse, 1982). An [s] noise has a relatively high pitch or bright timbre, which reflects the predominance of high-frequency components mostly above 4 kHz, whereas white noise has a somewhat lower (indefinite) pitch or darker timbre, which reflects its flat spectrum, and low-pass filtered noise has an even lower pitch. Therefore, subjects can be asked to make judgments about the relative pitch height of pairs of noises, and the term *pitch* will be used henceforth to refer to the subjective dimension along which the noises were compared. The prediction of the segregation hypothesis was, then, that the target noise should be perceived as having a lower pitch than the probe noise.

One possible complicating factor was that the acoustic context surrounding the target noise might have an influence on the perceived pitch of the noise for reasons that are unrelated to the hypothetical process of phone restoration. For example, the following context might interfere with a listener's memory for the noise pitch in specific ways, and both preceding and following context may exert auditory assimilation or contrast effects on perception of the noise spectrum. The memory problem was addressed by having the probe noise either precede or follow the speech in which the target noise was embedded. If the acoustic speech signal interfered with memory for the target noise, then this interference should be reduced when the probe noise comes first and subjects make a decision immediately upon hearing the target noise. The other problem (i.e., possible assimilation or contrast effects) required some control stimuli in which there was similar acoustic context but no incentive for fricative restoration. In Experiments 1 and 2, stimuli conducive to restoration of /s/ were compared with stimuli containing a /t/ instead. The phone replaced by the extraneous noise in the case of /t/ was silence (the stop closure interval), so that no acoustic energy was to be restored for perception of /t/. This seemed like an appropriate control condition to start with.

EXPERIMENT 1

According to the top-down completion hypothesis, a white noise burst that replaces an [s] noise in a sentence should be perceived as having the same pitch as an identical white noise burst occurring before or after the sentence (if we disregard the possibility of auditory context effects for the moment). According to the segregation hypothesis, however, the white noise in the sentence should sound lower in pitch (and also less loud, though this was not tested here) than a white noise probe, because the [s] frication has been perceptually subtracted from it, as it were, leaving a residual noise in which frequencies below 4 kHz predominate. It follows, then, that there must be some noise with relatively low-frequency predominance that, when employed as a probe, is perceived as matching the white noise target. The precise spectrum of that probe noise is not obvious, however, since it is not clear "how much" [s] noise the auditory system needs to subtract from the white noise inside the sentence to satisfy the needs of the speech processor. A strong version of the segregation hypothesis would predict that something close to the original [s] noise is subtracted, on the basis of expectations derived from the speaker's voice and knowledge about speech in general. However, a much weaker copy of that noise or a noise with a somewhat distorted spectrum might suffice for /s/ phoneme restoration.

To construct a noise probe that, according to the strong segregation hypothesis, should match the white noise target, the original [s] noise somehow needed to be subtracted physically from the white noise burst. Since arithmetic subtraction of the digitized [s] noise waveform from the white noise waveform does not have the desired spectral consequences (with random noise, waveform subtraction is essentially equivalent to waveform addition), a lowpass filtered noise (N) was instead created that, when the [s] noise waveform was added to it, resulted in a noise with an approximately flat spectrum (NS). N was the noise probe that was expected to match the NS target: If listeners restore [s] fully by segregation from NS, they should hear the intact speech accompanied by the residual extraneous noise N. (Note that the situation can be described either as NS replacing [s], or as N being added to [s]!) To simulate the more typical "added" condition, another noise was constructed by adding the waveforms of NS and [s] (NSS). According to the segregation hypothesis, NSS replacing [s] in a sentence should be perceived as [s] accompanied by NS. According to the top-down completion hypothesis, however, NS and NSS targets replacing [s] should be perceived as matching NS and NSS (not N and NS) probes, respectively.

In Experiment 1, then, three noise probes (N, NS, and NSS) were paired with two noise targets (NS and NSS). The target noise replacing [s] occurred either at the beginning or at the end of a trisyllabic word embedded in the constant carrier phrase "Say ... again," and the noise probe occurred either before or after the sentence. The temporal locations of the target and probe noises were varied to examine the role of auditory memory in the task. Because of limits on the size of the design, only one word of each type was employed: "seminar" and "happiness." The final /s/ of "happiness" is predicted by the preceding context, whereas the initial /s/ of "seminar" is not; however, to the extent that lexical constraints enhance phonemic restoration (Samuel, 1987), such effects were

not expected to play a role because of the presence of other acoustic cues to /s/ in the speech signal and because the same stimuli were repeated over and over in the course of the experiment. As to the problem of repetition itself, there is no evidence in the literature that repeated presentation reduces phonemic restoration (see Samuel, 1991; Warren & Obusek, 1971).

In addition to the two critical words, "seminar" and "happiness," two control words were used which had a /t/ instead of an /s/ in either initial or final position, in the same vocalic context: "telephone" and "cabinet." When a noise target was placed in these words, it occurred during the silent stop closure, so there was no acoustic material for listeners to restore. According to both hypotheses, the noise targets in these stimuli should be perceived veridically. A pilot study revealed a strong bias to perceive target noises as lower in pitch than probe noises. Such a bias should affect all stimuli equally, however. Therefore, the critical comparison in Experiment 1 was that between the test sentences containing /s/ and the control sentences containing /t/.

Method

Subjects. Ten volunteers from the Yale community were paid for their participation. All were native speakers of English who claimed to have normal hearing.

Stimuli. A female speaker produced the four sentences, "Say (seminar, happiness, telephone, cabinet) again," with neutral intonation in a sound-insulated booth. The utterances were recorded with high-quality equipment and were then digitized at a sampling rate of 20 kHz, with lowpass filtering at 9.8 kHz, but without high-frequency preemphasis. Using a digital waveform editor, the [s] noises of "seminar" and "happiness," 148 and 135 msec in duration, respectively, were excerpted and stored in separate computer files. Their average Fourier spectra were computed and inspected to determine the lower cutoff frequency of the main spectral energy. Each [s] noise was then converted into amplitude-modulated white noise by randomly reversing the polarity of sampling points with a probability of 0.5 (Schroeder, 1968). The resulting noises had exactly the same amplitude envelopes as the original [s] noises had, but a flat spectrum. Each of these noises was then passed through a digital lowpass filter whose cutoff frequency was set approximately where the energy of each [s] noise fell off (i.e., around 4 kHz). The attenuation level above the cutoff frequency was varied by trial and error until a noise resulted (N) that, when digitally added to the [s] noise waveform, yielded a noise with an approximately flat spectrum (NS). Before the noises were added together, the amplitude of the filtered noise was adjusted to match that of the natural [s]. A third noise (NSS) was created by adding the [s] waveform to NS with a temporal offset of a few samples, to avoid simply doubling the [s] sampling values in the mixture. The spectra of these noises derived from "seminar" and "happiness" are shown in Figure 1.[3]

Using the waveform editor, the NS and NSS noises were spliced back into the test sentences where the [s] noise had been. They were also spliced into the silent /t/ closures of "telephone" and "cabinet," which were extended to accommodate the noises. The original durations of the /t/ closures of "telephone" and "cabinet" were 67 and 76 msec, respectively. Since each closure coincided with a word boundary, it was reasonable to assume that extending its duration would have a negligible effect on /t/ perception.

Two stimulus sequences were recorded on audio tape. The first sequence served as practice and contained pairs of isolated noises. Pilot tests had suggested that without some training in the noise

matching task, subjects' performance was very poor. Pairings of N with NSS were omitted on this tape, because they were too easy to discriminate. Thus there were 3 identical (N-N, NS-NS, NSS-NSS) and 4 nonidentical (N-NS, NS-N, NS-NSS, NSS-NS) pairs for each of the two sets of noises, 14 pairs in all, which were recorded 12 times in random order, with interstimulus intervals (ISIs) of 750 msec within pairs and 3 sec between pairs. The second sequence, the main test, contained sentences paired with noise probes. There were five such combinations for each sentence: NS and NSS as targets paired with either NS or NSS as probes, and NS as target with N as probe. The probe occurred either before or after the sentence, with an ISI of 250 msec. Thus there were 40 possible stimuli altogether, which were recorded four times in random sequence, with ISIs of 3 sec.

Procedure. The subjects sat in a quiet room and listened over TDH-39 earphones. Their task was to write down one of three responses for each pair of noises: an up-pointing arrow if the second noise seemed higher in pitch than the first; a down-pointing arrow if it seemed lower; or an equal sign if the two noises seemed identical. The concept of noise pitch was first explained with reference to the phones [s] and [ʃ], [s] having a higher pitch than [ʃ]. It was emphasized that the differences to be detected would be small and that careful listening was necessary. The correct responses to the first 28 practice pairs of isolated noises were already filled in on the answer sheet, and subjects were requested merely to check them off while listening; the subjects then responded to the remaining noise pairs. Before the main test, subjects were told that the task remained the same, but that one of the noises in each pair would coincide with speech that was irrelevant and should be ignored. After a short break, the main test was repeated, so that each subject gave eight responses to each individual stimulus pair.

Results and Discussion

The results for the isolated noises are shown in Table 1. The subjects had little difficulty in discriminating the N and NS noises, but they often did not detect the difference between NS and NSS. Accuracy for identical pairs was intermediate. On the whole, these results are satisfactory, but the auditory similarity of NS and NSS must be kept in mind when interpreting the results of the main test.

These results are depicted in Figure 2. Each of the five panels compares the results for control (/t/) and test (/s/) sentences for one of the five probe–target combinations. The four functions in each panel represent the two possible temporal orders of probe and target (P–T or T–P) for each of the two possible locations in the word (initial or final). The quantity plotted on the ordinate is the difference between the percentages of "probe higher than target" and "probe lower than target" judgments. These scores thus range from −100 to 100, with 0 representing perceived identity. ("Equal" judgments were not considered; this way of treating the data is equivalent to assigning weights of −1, 0, and 1 to the three types of responses.) A repeated measures analysis of variance (ANOVA) was conducted on the individual subjects' scores, with the N/NS condition omitted to balance the design. There were six crossed within-subject factors in the ANOVA: condition (/t/ vs. /s/), probe (NS vs. NSS), target (NS vs. NSS), location (initial vs. final), order (P–T vs. T–P), and repetitions (first vs. second presentation of the main test).

The most striking feature of Figure 2, and also the most important result, is that in 19 out of 20 comparisons (in-
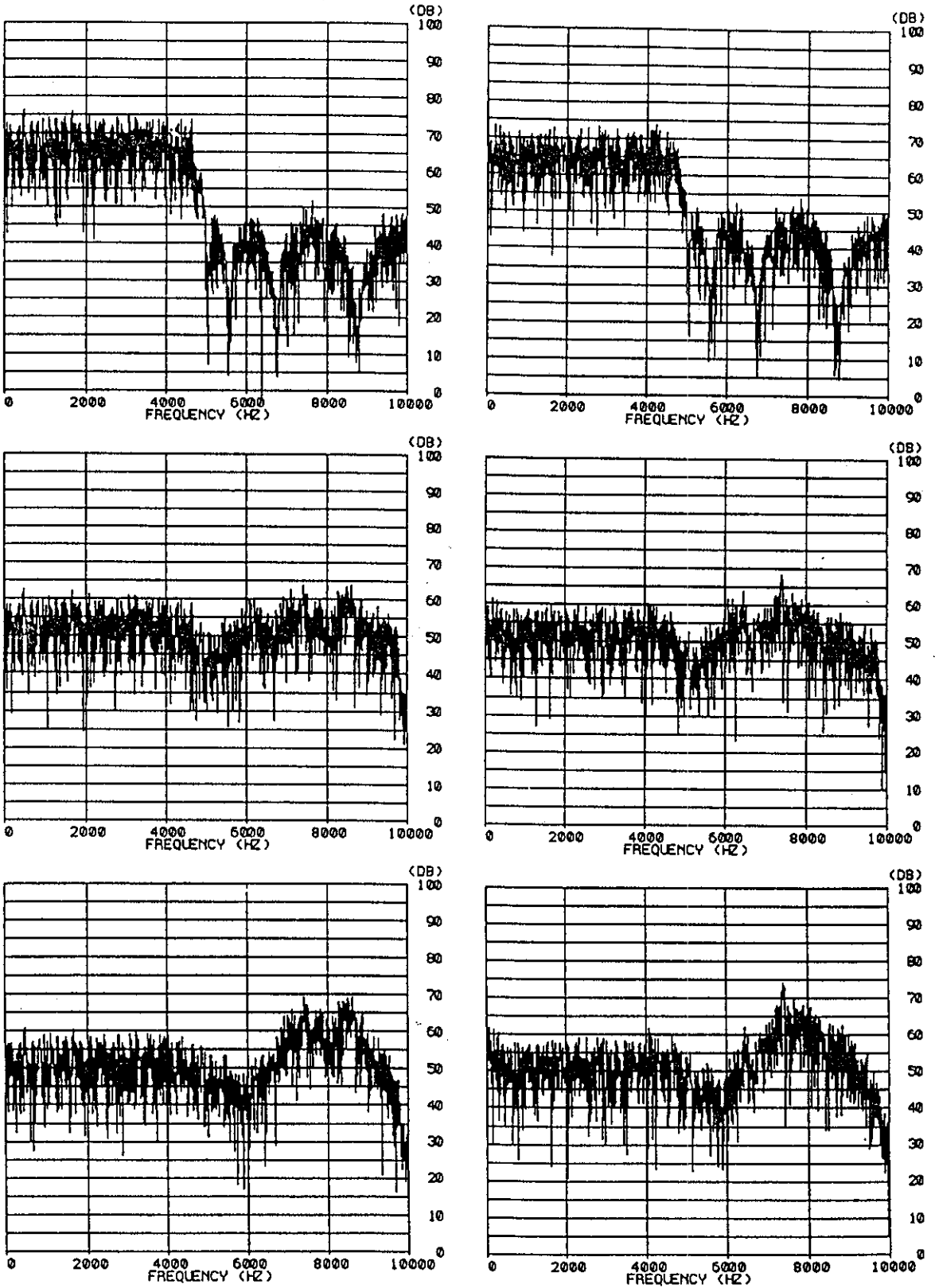
Figure 1. Fourier spectra (without high-frequency preemphasis) of the N (top), NS (middle), and NSS (bottom) noises. The spectra for "seminar" are on the left; those for "happiness" are on the right.

Table 1
Average Response Percentages for Isolated Noise Pairs in Experiment 1
(Second Noise Lower Than/Equal to/Higher Than First Noise)

First Noise	Second Noise																	
	Initial Type									Final Type								
	N			NS			NSS			N			NS			NSS		
N	10	82	8	0	2	98				10	88	2	8	3	89			
NS	98	0	2	13	76	12	13	40	48	97	3	0	9	88	3	5	44	51
NSS				61	29	10	8	78	14				38	56	6	5	85	10

dividual lines in the figure) the test (/s/) sentence probes received higher scores overall than the control (/t/) sentence probes did. The condition main effect was highly significant [$F(1,9) = 23.83, p < .001$]. This means that, relative to the control baseline, the targets in the test sentences were perceived as lower in pitch than the targets in the control sentences. This result supports the segregation hypothesis and suggests that the restored [s] noise was perceptually subtracted from the noise in the sentence.

The importance of including the control baseline is underscored by the presence of a strong bias to judge probes as higher in pitch than targets: Seven out of eight scores for physically identical probe–target combinations (NS/NS, NSS/NSS) in control (/t/) sentences were well above zero (see the second and fifth panels in Figure 2).

Even though the NS and NSS noises were not easy to discriminate in isolation, there were significant main effects of probe [$F(1,9) = 18.35, p < .003$] and of target [$F(1,9) = 9.90, p < .02$], which show that the subjects could distinguish the two noises to some extent even in the main test. Target noises seemed to be more difficult to distinguish than probe noises, because of the surrounding sentence context.[4]

If auditory segregation of [s] from the target noise had been complete, the N probe should have been perceived as matching the NS target in the test sentences; however, the N probe was clearly perceived as still much lower in pitch (first panel in Figure 2). The segregation hypothesis also predicted that the NS probe should be perceived as matching the NSS target in the test sentences, but instead the probe was perceived as higher in pitch, contrary to the actual physical difference (fourth panel in Figure 2). This was due to the general bias to perceive probe noises as higher in pitch than target noises. Therefore, these results need to be compared with the scores for physically identical probe–target pairs (NS/NS and NSS/NSS) in control sentences; these seem comparable, being subject to the same bias.
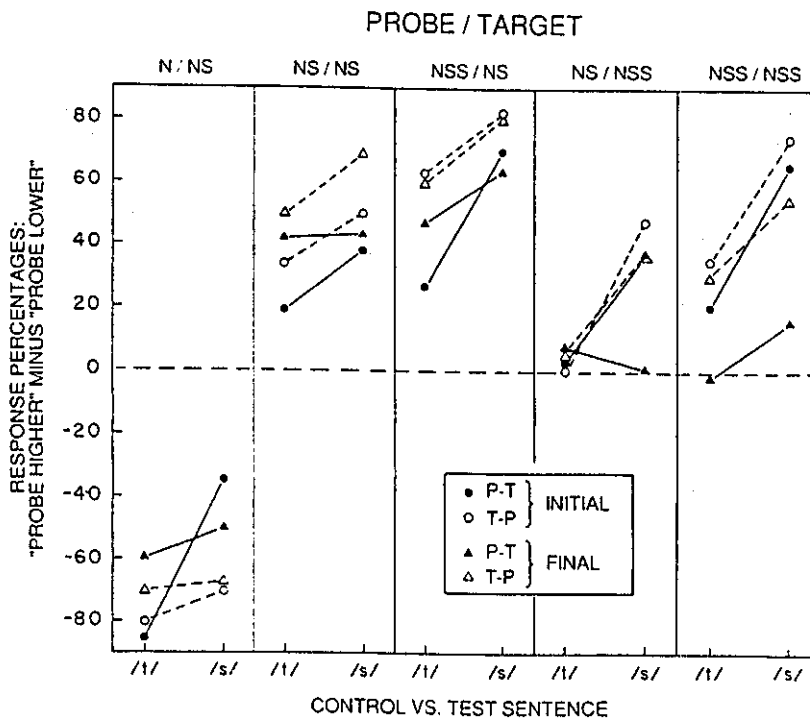
PROBE / TARGET

Figure 2. Results of Experiment 1 (main test). Each panel shows one particular probe–target combination. P–T, probe followed by target; T–P, target followed by probe. Initial = word-initial fricative; final = word-final fricative. The rising slope of the lines is of main interest.

Thus, although the predictions of the strong segregation hypothesis are not upheld in one condition (N probe/NS target, NS targets corresponding to the "replaced" items in Samuel's design), they are consistent with the results for another condition (NS probe/NSS target, NSS targets corresponding to Samuel's "added" items). Note, however, that the NS target can also be regarded as resulting from the addition of an extraneous lowpass filtered noise N to [s], whereas the NSS target could have resulted from replacing the [s] noise with a spectrally similar noise NSS. Ultimately, therefore, the distinction between "added" and "replaced" becomes meaningless, and what takes its place is the continuous dimension of spectral similarity between the target noise and the original [s] noise (or, rather, the [s] noise that a listener might expect to occur, since listeners never actually heard the original [s] noise in this experiment, though they heard an /s/ in the word "say" at the beginning of each sentence). It seems, then, that [s] restoration is more complete when the target noise is spectrally similar to an [s] noise (NSS) than when it is not (NS). Listeners seem to perceptually subtract more from an NSS target than from an NS target. In support of this suggestion, the author observed informally as a listener that, even though /s/ restoration clearly occurred in both conditions, the /s/ in the NS target condition seemed less natural than that in the NSS target condition; it had a "lisping" quality. Phone restoration thus seemed to be incomplete in the NS target condition, even though the original [s] noise was physically present in the target noise. This observation will be discussed further in the General Discussion.

EXPERIMENT 2

The results of Experiment 1 were highly suggestive of auditory segregation, but there were some irregularities in the results, and a replication seemed desirable. Experiment 2 provided a partial replication, with probes of a finer grain. Since so few data points fell close to zero in Experiment 1 (see Figure 2), it was of interest to determine more precisely what kind of probe was judged to match the target, particularly for the NS target condition. This was achieved by constructing a continuum of five probe noises between N and NS, and using two of them as targets. Because of the increase in probe–target combinations, only the initial-location stimuli ("seminar" and "telephone") were used, which had yielded more consistent data in Experiment 1.

One additional change in Experiment 2 concerned the instructions for the repetition of the main test. Whereas in the first presentation, subjects were asked to ignore the speech, as in Experiment 1, during the second presentation they were required to identify the critical word in the sentence by writing down its initial letter after making a judgment about the noise pair. This additional requirement was intended to increase subjects' attention to the speech. The question was whether this would affect the extent of auditory segregation. For example, it could be

that phone restoration in Experiment 1 was incomplete because attention was directed away from the speech, toward the extraneous noise, even though this seems unlikely.

Method
Subjects. Ten paid subjects from the same general population participated.

Stimuli. The stimuli, derived from "seminar" and "telephone," were the same as in Experiment 1, with the exception of the following changes in probe and target noises. The N and NS noises were renamed N1 and N5, respectively, and three additional noises with intermediate spectra (N2, N3, N4) were created by adding the digitized waveforms (matched for overall amplitude) of N (the lowpass filtered noise) and the original [s] noise with the following weights: 0.875/0.125, 0.75/0.25, and 0.625/0.375 (corresponding approximately to N/[s] ratios of 17, 9.5, and 4.5 dB). N5, of course, represented the addition of N and [s] with equal weights of 0.5 (N/[s] ratio of 0 dB). These noises thus constituted a series of increasing pitch. All five noises were used as probes; however, only N3 and N5 were used as targets. The asymmetry between probe and target noises (N1 was not used as a target) anticipated (but possibly also encouraged) the bias to judge probes as higher in pitch than targets.

As in Experiment 1, two test sequences were recorded. The first, for practice, contained pairs of isolated noises with ISIs of 750 msec within pairs and 3 sec between trials. N3 and N5 occurred in all possible pairings with each other and with all other noises: 7 nonidentical pairs (in two possible orders) and 2 identical pairs (repeated once), a total of 18 stimulus pairs. Six different randomizations of these 18 pairs were recorded. The second test sequence, the main test, contained five randomizations of 40 noise pairs. One noise in each pair (the target) was buried in one of the two speech contexts; the other noise (the probe) preceded or followed the speech with an ISI of 250 msec. The intertrial interval was 3 sec. The 40 pairs resulted from five probes paired with two targets in two speech contexts in two temporal orders. The tape was played twice.

Procedure. The procedure and instructions were exactly the same as in Experiment 1 except that, for the second presentation of the main test, subjects were asked to write next to their judgment of the noises the letter S or the letter T, depending on whether the speaker on the tape had said "seminar" or "telephone."

Results and Discussion
The results for the isolated noise pairs are shown in Table 2. It can be seen that the subjects, while far from perfect, were clearly sensitive to the spectral differences among the noises, and their accuracy was quite satisfactory for the purposes of this experiment.

The results of the main test, pooled over the two repetitions, are shown in Figure 3. The response measure is the same as in Figure 2; the higher the score, the higher in pitch the probe was judged relative to the target, with zero representing perceived equality. On the abscissa we have now a continuum of probe noises, with separate panels for preceding and following probes. It is evident that subjects' responses increased as an orderly function of probe spectrum, as in the pretest [$F(4,36) = 78.11$, $p < .0001$, for the probe main effect]. Probe discrimination seemed slightly better when the probe followed than when it preceded the speech, as indicated by the steeper functions in the right-hand panel of Figure 3 [$F(4,36) = 5.74$, $p < .002$, for the probe × order interaction]. Clearly, the two targets were also discriminated, with higher ratings for probes paired with N3 than for probes

Table 2
Average Response Percentages for Isolated Noise Pairs in Experiment 2
(Second Noise Lower Than/Equal to/Higher Than First Noise)

First Noise	Second Noise														
	N1			N2			N3			N4			N5		
N1							2	10	88				0	0	100
N2							10	36	54				2	0	98
N3	98	0	2	78	16	6	12	70	18	10	46	44	0	24	76
N4							32	62	6				6	30	64
N5	98	0	2	98	2	0	80	18	2	60	14	26	4	88	8

paired with N5 [$F(1,9) = 34.31, p < .0003$, for the target main effect]. This difference was somewhat more pronounced when the probe preceded the target—that is, when the target did not have to be held in memory [$F(1,9) = 13.48, p < .006$, for the target × order interaction]. Most importantly, however, there was a consistent difference between the /s/ and /t/ conditions: For all probe-target combinations, the probe was judged higher (hence, the target was judged lower) in pitch when [s] was to be restored, in agreement with the segregation hypothesis [$F(1,9) = 83.27, p < .0001$, for the condition main effect]. The difference was somewhat larger when the probe preceded the target [$F(1,9) = 12.49, p < .007$, for the order × condition interaction]. There were also significant target × probe [$F(4,36) = 3.76, p < .02$] and condition × target × probe interactions [$F(4,36) = 4.27, p < .007$], reflecting differences in the shapes of the response curves that were due to floor and ceiling effects. (No data transformation was applied.) The ANOVA further revealed a repetitions × order interaction [$F(1,9) = 7.64, p < .03$], which was due to an increase in the order

effect in the second half of the main test, and a marginally significant four-way interaction. On the whole, however, instructions to pay attention to the speech in the repetition of the main test had little effect (nor did practice or fatigue, for that matter).

Thus the results of this experiment again lend support to the segregation hypothesis, even for a target noise (N3) that was rather dissimilar to [s]. As in Experiment 1, however, the apparent phone restoration was far from complete. If it had been, the N5 target should have been judged as matching the N1 probe; in fact, it was judged closer to N3 or perhaps N2 (filled circles in Figure 3). A large part of that shift, moreover, must be attributed to the general bias to perceive probes as higher in pitch, which was again present: In three out of four pairings of probes with identical targets replacing /t/, the probe was judged considerably higher in pitch than the target (N3/N3 and N5/N5 probe/target pairs, open symbols in Figure 3). If this bias is taken into account, the N5 target seemed most similar to the N3 or N4 probe. Thus the [s] segregation effect, though consistently present, was relatively small. Clearly, listeners did not extract a full [s] from the target noise.

EXPERIMENT 3

The initial two experiments have provided evidence of partial segregation of an [s] that has been replaced with a different noise. The question must be raised now, however, whether this apparent segregation was induced by top-down expectations (as in the standard phoneme restoration paradigm) or by residual cues to the "missing" phone in the acoustic signal. Even though the formant transitions in adjacent vocalic segments were not highly distinctive between /s/ and /t/, and even though cues to fricative manner were weak, the presence of the /t/ release burst plus aspiration provided salient information for the presence of a /t/ in the control sentences. The difference between the /s/ and /t/ conditions, on which the conclusions of Experiments 1 and 2 rest, could thus be due to an inhibition of [s] restoration by local acoustic context (rather than by lexical constraints) in the control condition, or, what is worse, it could even reflect some direct effect of acoustic environment on the perceived pitch of the target noise.

If phone restoration were facilitated or inhibited by acoustic environment, this would be an interesting finding in itself. However, a vital concern of current models of word recognition (e.g., Marslen-Wilson, 1987; McClelland &
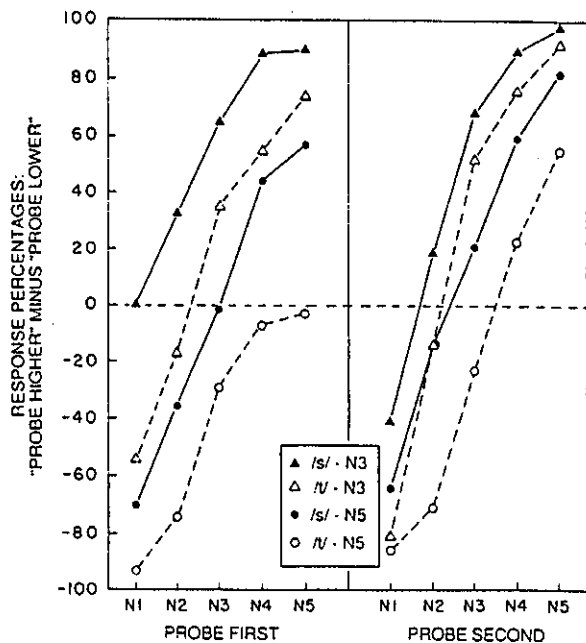
Figure 3. Results of Experiment 2 (main test). N3 and N5 (legend) are the target noises; /s/ and /t/ are the phonemes to be restored.

Elman, 1986) is whether lexical top–down effects can penetrate to relatively early levels of speech processing. Elman and McClelland (1988) have argued that they can, after showing that a lexically disambiguated (they call it "restored") fricative can influence the perception of a following stop consonant. (See also Mann & Repp, 1981, who reported essentially the same finding for spontaneous fricative disambiguation.) *Phoneme* restoration was lexically driven in the classic experiments in which all acoustic cues were removed. Experiment 3 examined whether *phone* restoration and auditory segregation would be observed when differences in bottom-up contextual cues were eliminated. This was accomplished by making the acoustic environment of the target noise in "telephone" identical to that in "seminar," favoring either /s/ or /t/ in either context. If bottom-up factors were solely responsible for the segregation effect, the "telephone"/"seminar" main effect should disappear under these conditions.

A second question addressed in Experiment 3 concerned the strong bias to perceive target noises as lower in pitch than probe noises. One possible origin of this tendency is that the /t/ environment, because of its relative similarity to that of /s/ and because of the artificial lengthening of the stop closure interval, actually encouraged perception of /s/ (only less so than the /s/ environment), even though the /t/ was aspirated and a nonword was the result ("selephone" or "stelephone"). In Experiment 3, therefore, yet another control condition was employed, in which the target noise not only replaced a segment (a /p/ closure) whose acoustical context was dissimilar from that of either /s/ or /t/, but also was followed by a segment (/r/) that phonotactically barred the occurrence of an immediately preceding /s/. If the bias were absent in that condition, then it would have to be attributed to /s/ restoration in the /t/ condition.

Method

Subjects. Ten new subjects from the same general pool participated.

Materials. A new set of six utterances was recorded by a different female speaker. The carrier phrase was once more "Say ... again," and the embedded items were "seminar," "teminar," "selephone," "telephone," "president," and "tresident." The first four items thus varied lexical status (word or nonword) and local acoustic information (/s/ or /t/ cues) independently, whereas the last two items introduced a phonotactic bias against /s/ while also varying local information and lexical status.

Target and probe noises (N1–N5) were constructed exactly as in Experiment 2, using as the starting point a 128-msec [s] noise excerpted from another utterance containing "selephone." Again, only N3 and N5 were used as targets, whereas all five noises were used as probes. The pretest contained 12 random sequences of 18 pairs of isolated noises (twice as many as in Experiment 2, to provide additional practice). The main test contained five blocks of 60 target–probe pairs (6 items × 2 targets × 5 probes). There were two important differences with respect to Experiment 2: First, to reduce the number of stimulus pairs, the probe noise always followed the target. Second, in the items originally containing a stop consonant, the target noise replaced the closure and the release burst; the aspiration following the release burst was left intact but provided a much weaker cue to stop consonant manner.

Procedure. As in Experiment 2, the subjects listened first to the pretest, with correct answers provided for the first two blocks. The main test was presented twice. During the second presentation, the subjects were required to write down, in addition to their judgment of the noises, the initial consonant of the item containing the target. They were told that the speaker sometimes mispronounced the words. The two possible choices (S/T for "-eminar" and "-elephone," and P/T for "-resident") were listed on top of the answer sheet.

Results and Discussion

The pretest results, shown in Table 3, resemble those obtained in Experiment 2. Clearly, the subjects were able to discriminate the noises quite well in isolation.

The results of the main test are presented in Table 4. The response functions (rows in the table) were steep; clearly, the probe noises were discriminated well by the subjects [$F(4,36) = 165.96, p < .0001$, for the probe main effect]. There was also a reliable difference between the two target conditions, though it was small in size [$F(1,9) = 24.31, p < .0009$]. However, there were no other interesting differences in the data. Several other significant effects in the ANOVA were all due to an unexplained anomaly: a reversal of the target effect for the "telephone" items (see Table 4, N3 probe).

The prediction of the segregation hypothesis was that scores would be higher for /s/ items than for /t/ items, and in particular higher for "seminar" than for "telephone," as they were in Experiment 2. However, these differences were small and nonsignificant, even slightly in the opposite direction. Thus the results did not replicate the principal finding of Experiments 1 and 2—namely, that [s] restoration affected perception of the extraneous noise, nor did they demonstrate any systematic effect of local acoustic environment. They did show, however, the by now familiar bias to perceive probes as higher in pitch than targets, though only for N5 targets, which were perceived as similar to N3 probes. N3 targets, on the other hand, were also perceived as most similar to N3 probes.

What phonemes did the subjects actually restore? In Experiments 1 and 2, correct phoneme restoration was taken

Table 3
Average Response Percentages for Isolated Noise Pairs in Experiment 3
(Second Noise Lower Than/Equal to/Higher Than First Noise)

First Noise	Second Noise														
	N1			N2			N3			N4			N5		
N1							4	16	80				2	0	98
N2							8	52	40				2	2	96
N3	98	2	0	88	8	4	16	77	7	2	46	52	4	6	90
N4							70	24	6				16	42	42
N5	100	0	0	96	2	2	82	12	6	16	64	20	2	80	18

Table 4
Average Scores (Percent Judgments of "Probe Higher" Minus Percent Judgments of "Probe Lower")
in the Main Test of Experiment 3, and Average Percentages of Consonant Identification Responses

Item	Target	Probe					Percent Responses		
		N1	N2	N3	N4	N5	S	T	P
"seminar"	N3	−85	−47	22	68	84	77	23	
	N5	−84	−57	−1	52	87	97	3	
"teminar"	N3	−84	−42	14	69	89	36	64	
	N5	−93	−65	−23	57	81	60	40	
"selephone"	N3	−78	−39	11	66	91	89	11	
	N5	−88	−66	−11	61	89	96	4	
"telephone"	N3	−69	−46	1	75	86	12	88	
	N5	−85	−50	17	74	85	33	81	
"president"	N3	−86	−42	15	69	90		14	86
	N5	−84	−64	−8	45	74		21	79
"tresident"	N3	−69	−41	6	56	91		42	58
	N5	−86	−57	−12	34	83		54	46

for granted because lexicality and acoustic cues coincided. In the present experiment, these two factors were dissociated. Were subjects guided by lexicality or by acoustic cues or by both? A tally of subjects' initial consonant responses in the second presentation of the main test provided an answer. The average percentages of consonant responses are shown in the last three columns of Table 4. Each subject gave 25 responses to each item. There was considerable variability in subjects' identification of the items. It should be noted first, however, that the two items that had also been used in the previous experiments, "seminar" and "telephone," were generally identified correctly (i.e., as intended by the speaker of the materials). Although a substantial number of "errors" did occur, they are not sufficient to explain the total absence of a segregation effect in the present experiment. It can further be seen in Table 4 that the items "selephone" and "president" were generally identified correctly. Not so for "teminar" and "tresident," however, which tended to be identified as "seminar" and "president," respectively, about half of the time. In addition to these differences among items, which apparently reflect both acoustic and lexical factors, an influence of the target noise on consonant identification can be seen: In the first four items, where S and T were the response choices, N3 (the lower pitched noise) favored T, while N5 (the higher pitched noise) favored S. In the last two items, where P and T were the choices, N3 favored P and N5 favored T. Since the noisy acoustic correlates of /p/, /t/, and /s/ form a series of increasing pitch, the influence of the target noise spectrum on phoneme restoration may represent an instance of "bottom-up confirmation" (Samuel, 1981b).

The effects just described were statistically reliable. A $2 \times 2 \times 2$ ANOVA was conducted on the S response percentages to the the first four items ("s/teminar," "s/telephone"), with the factors context ("-eminar" vs. "-elephone"), cues (/s/ vs. /t/), and target (N3 vs. N5). The main effects of cues and of target were highly significant [$F(1,9) = 40.54$, $p < .0002$; $F(1,9) = 61.12$, $p < .0001$]. The context $\times$ cues interaction was likewise significant [$F(1,9) = 31.88$, $p < .0004$] and probably reflects an effect of lexicality (words vs. nonwords). In addition, there was a context main effect [$F(1,9) = 12.39$, $p = .007$], due to more S responses to "s/teminar" than to "s/telephone," and a context $\times$ target interaction [$F(1,9) = 6.08$, $p < .04$], due to a larger target effect for "s/teminar" than for "s/telephone." A separate 2×2 ANOVA on the "s/tresident" items revealed significant cues and target main effects [$F(1,9) = 9.99$, $p < .02$; $F(1,9) = 23.36$, $p < .001$]. Lexicality was confounded with cues here.

In summary, these data show that the identity of the restored phoneme is influenced by both acoustic and lexical factors. The acoustic determinants in particular are very clear: Both the residual cues in the signal (despite elimination of the stop release bursts) and the spectral properties of the target noise had effects. The effect of lexicality seems less systematic because of the accurate identification of "selephone."[5] In the other two comparisons, the real words ("seminar," "president") were identified more accurately than the nonwords ("teminar," "tresident"), and so was the real word "telephone," which shared the initial consonant with these nonwords. Since all items became thoroughly familiar to the subjects in the course of the experiment, lexical biases were presumably attenuated relative to a more natural situation.

EXPERIMENT 4

Even though the subjects generally identified the critical test items as intended, the segregation hypothesis was not supported in Experiment 3. The subjects also had some difficulty distinguishing the target noises in the main test. One disturbing possibility is that the subjects may have adopted a strategy of essentially ignoring the target noise and making their noise comparison judgments on the basis of the probe noise alone, responding "higher" when its pitch was high and "lower" when its pitch was low. This strategy may have been encouraged by the fact that the probe noise always followed the sentence, so that the target had to be held in auditory memory for comparison, whereas the probe noise was freshly perceived. In

Experiment 4, therefore, the design of the earlier experiments, with probes both preceding and following the targets, was restored. To reduce the number of stimuli, the N1 and N5 probes, which provided little information, were dropped. The replication experiment was also desirable because of the anomalous data for the "telephone" stimuli in Experiment 3, and because of the restrictive two-alternative choice task in consonant identification. In Experiment 4, the subjects were free to write down any initial consonant they heard during the second half of the main test.

Method

Subjects. Eleven new subjects were recruited. Three subjects' data were dropped, however, because of poor performance in the pretest or the main test, which left only 8 subjects.

Materials. The stimuli were the same as in Experiment 3. A new main test was constructed in which the probe noise either preceded or followed the sentence. To compensate for the doubling in the number of trials, only N2, N3, and N4 were used as probes. This resulted in a test comprising five blocks of 72 target-probe pairs (6 items × 2 targets × 3 probes × 2 orders).

Procedure. The procedure was the same as that in Experiment 3, except that, in the second presentation of the main test, the choices for consonant identification were not listed, and subjects were free to write down whatever consonant they heard.

Results and Discussion

The pretest of isolated noises was the same as in Experiment 3, and the results for it are shown in Table 5. Performance was comparable to that in Experiment 3.

The results of the main test are shown in Table 6. It is evident that the subjects performed very poorly when the probe noises preceded the target, in contrast with their performance in Experiment 2, for reasons that are not quite clear. A separate ANOVA on this half of the data showed neither a significant main effect of probe [$F(2,14) = 1.13$, $p > .35$] nor one of target [$F(1,7) = 1.32$, $p > .28$]. The only significant effects were a target × repetitions interaction [$F(1,7) = 8.32$, $p < .03$], suggesting that the subjects began to discriminate the target noises in the second half of the test, and a marginal higher level interaction also involving repetitions. Thus, this part of the data was rather useless.

Fortunately, the results for following probes look better, at least as far as subjects' sensitivity to the probe noises is concerned. The ANOVA on this part of the data (in which the six items were treated for convenience as representing a 3×2 context × cues design, even though the cues in the "p/tresident" pair were different) showed a highly significant main effect of probe [$F(2,14) = 18.27$, $p < .0002$], and the subjects' response functions (truncated here) were comparable to those obtained in Experiments 2 and 3. However, there were no other significant effects. In particular, the target main effect was nonsignificant [$F(1,7) = 1.36$, $p > .28$], indicating that the subjects were unable to discriminate the target noises, and the cues and context main effects, as well as their interaction, were nonsignificant, indicating neither acoustic nor lexical influences on perception of the target noise as such.

In a further attempt to detect any systematic tendencies in the data, an ANOVA was performed on the combined data of Experiments 3 (omitting data for the N1 and N5 probes) and 4 (data for following probes only). Because

Table 5
Average Response Percentages for Isolated Noise Pairs in Experiment 4
(Second Noise Lower Than/Equal to/Higher Than First Noise)

First Noise	Second Noise														
	N1			N2			N3			N4			N5		
N1							10	15	75				3	0	97
N2							5	48	47				0	5	95
N3	100	0	0	92	3	5	14	67	19	5	33	62	3	5	92
N4							59	28	13				3	35	62
N5	97	0	3	95	0	5	67	23	10	25	67	8	4	61	35

Table 6
Average Scores (Percent Judgments of "Probe Higher" Minus Percent Judgments of "Probe Lower") in the Main Test of Experiment 4

Item	Target	Probe First			Probe Second		
		N2	N3	N4	N2	N3	N4
"seminar"	N3	−21	1	6	−45	−1	53
	N5	3	5	13	−45	−8	38
"teminar"	N3	−16	−15	5	−21	−1	38
	N5	−13	3	14	−53	8	46
"selephone"	N3	−15	−14	−6	−33	−11	49
	N5	1	−6	26	−39	−6	48
"telephone"	N3	−35	−4	9	−44	24	44
	N5	−11	6	3	−51	18	34
"president"	N3	−8	−1	8	−40	6	53
	N5	−6	5	16	−49	11	51
"tresident"	N3	−10	8	6	−24	−1	59
	N5	−8	14	8	−43	11	31

Table 7
Average Percentages of Consonant Identification (Open Choice) in Experiment 4

Item	Target	Response						
		S(C)	T	P	F	TH	H	Other*
"seminar"	N3	91	0	0	8	0	0	0
	N5	98	0	0	1	0	0	0
"teminar"	N3	85	0	0	7	0	6	1
	N5	95	0	0	3	0	2	0
"selephone"	N3	61	17	0	7	16	0	0
	N5	68	15	0	7	10	0	0
"telephone"	N3	30	40	0	9	14	5	2
	N5	35	39	0	8	15	3	1
"president"	N3	0	0	43	57	0	0	0
	N5	1	0	41	58	0	0	0
"tresident"	N3	0	0	40	59	1	1	0
	N5	0	0	40	59	0	0	0

*A few each of ST, STH, SH, CH. Rows may not add up to 100% because of rounding error.

of the increased statistical power, a number of effects were significant in this analysis. Thus there were the expected main effects of probe $[F(2,34) = 177.25, p < .0001]$ and of target $[F(1,17) = 18.08, p < .0006]$, as well as a number of significant interactions. The only interaction of potential interest, however, between context, cues, and probe $[F(4,34) = 4.50, p < .003]$, was due to a complex and essentially uninterpretable pattern of results. The main prediction of the segregation hypothesis, that scores should be higher in items with /s/ cues than in items with /t/ cues, was not supported, and neither the main effect of context or of cues nor their interaction was significant.

Table 7 shows how the subjects in Experiment 4 identified the restored initial consonants, without any restraint on their choices and without any information on how the words might have been "mispronounced" by the speaker. The data must be viewed with some caution, because listeners differed substantially in their reactions to these stimuli. (One subject, for example, identified all items as intended by the speaker.) What the results show, however, is that responses other than the expected ones (S, T, P) were quite frequent, particularly F (for "-resident") and TH (for "-elephone"); H responses to stimuli originally starting with /t/ may have reflected listeners' perceptions of the aspiration noise remaining in the signal. There are some striking differences from the forced-choice results of Experiment 3, in that the initial consonants of "-eminar" and "-resident" were *never* identified as T, and even "telephone" received fewer T responses than previously. Clearly, restoration of stop consonants was not encouraged by the target noise, and fricative responses predominated. There was no clear effect of residual acoustic cues here, except in "-elephone," where T responses were more frequent to "telephone" than to "selephone"; this could also have been a lexical bias, however. S responses were more frequent to "seminar" than to "selephone," whereas T responses were more frequent to "telephone" than to "teminar," which may be an effect of lexicality. There was also a small but consistent effect of target noise; as in Experiment 3, the higher pitched noise (N5) encouraged /s/ restoration. Because of the high intersubject variability, however, no statistical analysis was conducted on these data.

As for the two items shared with Experiment 2, the results show very clearly that /s/ restoration did occur in "seminar," whereas listeners were less certain about the initial consonant of "telephone" but reported /s/ only one third of the time. Nevertheless, as in Experiment 3, there was no significant difference in subjects' perception of the extraneous noise in these items. Taken together, the data of Experiments 3 and 4 raise the possibility that the results of Experiments 1 and 2, which had been interpreted as supporting the segregation hypothesis, were in fact not caused by [s] segregation. The "telephone" stimuli of Experiments 1 and 2 differed from those of Experiment 3 in that they retained the /t/ release burst in the acoustic signal. This high-frequency burst may have affected subjects' perception of the target noise in an assimilative fashion, making the noise seem higher in pitch, or it may itself have been treated as part of the noise to be judged. The results of Experiments 3 and 4 offer no support for the segregation hypothesis; by suggesting that phone restoration (to the extent that it occurs at all) leaves perception of the extraneous noise unaffected, they seem to favor the top-down completion hypothesis.

In addition, Experiments 3 and 4 show that the general bias to perceive the target noise as lower in pitch than the probe noise was not due to [s] restoration in stimuli meant to induce restoration of some other phone. The bias was equally present in "-resident" stimuli, which never led to [s] restoration. Although /f/ percepts were common in these stimuli, [f] frication is lower in both pitch and intensity than [s], which should have been reflected in perception of the target noise if segregation had been operative. Instead, the results suggest that the bias represents a general effect of acoustic context on timbre perception: The perception of high-frequency spectral components may be selectively impaired when a noise is embedded in other sounds.

EXPERIMENT 5

In view of the negative results of Experiments 3 and 4, it seemed desirable to obtain further data, using different stimuli. One thing all experiments so far had in common was the repetition of a limited set of speech stimuli to ob-

tain stable results for each. Although there was reason to believe that extensive exposure and familiarity with the stimuli would not reduce the strength of phone (or phoneme) restoration, the possibility cannot be dismissed that the desired phenomenon was perhaps not elicited in its full strength. In the typical phoneme restoration experiment (as exemplified by Samuel's work), many different lexical items are included. Experiment 5 followed this design.

The crucial comparison in Experiment 5 was that between stimuli favoring restoration of /s/ and stimuli favoring restoration of /ʃ/. The corresponding fricative noises are known to differ in pitch, with [ʃ] being lower than [s] on this dimension. Consequently, the segregation hypothesis would be supported if target noises were judged to be lower in pitch when [s] is restored than when [ʃ] is restored.

To induce restoration of these fricative phones, lexical, acoustic, and phonotactic factors were manipulated in the stimuli. Half the stimuli contained /s/ cues and half contained /ʃ/ cues in the phonetic context. Within each of these sets, some stimuli were lexically ambiguous (i.e., they formed a word if the initial consonant was either /s/ or /ʃ/), some were lexically unique (i.e., only one of the two fricatives made a word), and some of these had an additional phonotactic restriction against nonword percepts (the second consonant was /l/ or /r/; word-initial /sl/ and /ʃr/ are common in English, but /sr/ and /ʃl/ are not). If segregation were operative, it should be evident most strongly in this last group of stimuli, which most strongly induce restoration of the intended fricative.

Method

Subjects. Ten new subjects from the same general population were tested.

Stimuli. Eighty familiar, monosyllabic words were selected from a dictionary; they are shown in Table 8. Half of them began with /ʃ/ and half with /s/. As can be seen in the table, half of the words formed minimal pairs distinguished only by the initial consonant (columns 1 and 2). The remaining words had no lexical counterpart beginning with the other fricative (columns 3 and 4). In half of these words, the fricative was followed by a vowel, so that the nonword resulting from substitution of the other fricative was phonotactically regular (e.g., "sade," "shafe"). In the other half, the following consonant was a liquid, /r/ after /ʃ/ and /l/ after /s/, so that nonwords resulting from fricative substitution would be phonotactically irregular (e.g., "srewd," "shlash"). This was expected to introduce a strong bias against perception of such nonwords in the experiment.

The 80 words were spoken by a male speaker in the carrier phrase "Say . . . again," in random order and with neutral intonation. The recordings were digitized at 20 kHz without preemphasis, lowpass filtered at 9.8 kHz, and edited to remove the fricative noise at the onset of each test word. The durations of these original noises ranged from 120 to 226 msec. One of the noises (from "sack," 178 msec long) served as the basis for generating the target and probe noises, which were slightly different from those used in the previous experiments. The noise was first converted into signal-correlated noise. This envelope-matched broadband noise was used as the target noise in all stimuli, in the hope that the diversity of the materials would prevent the subjects from becoming conscious of the fact that the target noise remained constant. Four additional noises were generated by lowpass filtering the target noise with a cutoff frequency

Table 8
Stimulus Words Used in Experiment 5

shack	sack	shade	safe
shag	sag	shaft	salt
shame	same	shape	sand
shave	save	share	seal
sheep	seep	sharp	send
sheet	seat	shawl	silk
sheik	seek	shirt	size
shelf	self	shout	soft
shell	sell	shove	soil
shift	sift	shut	surf
shin	sin	shred	slam
shine	sign	shrewd	slash
ship	sip	shriek	slave
shock	sock	shrill	sleep
shoe	sue	shrimp	slide
shoot	suit	shrine	sling
shore	sore	shrink	slip
short	sort	shroud	slow
show	sew	shrub	slot
shy	sigh	shrug	slur

of 5 kHz, using high-frequency stop-band attenuation levels of 12, 9, 6, and 3 dB. These probe noises, together with the original broadband noise, which was also used as a probe, formed a continuum of increasing pitch. They will again be referred to with the labels N1 to N5, N1 standing for the probe noise with the lowest pitch, and N5 standing for the original broadband noise that was also used as the target noise. In this experiment, therefore, no probe noises were actually higher in pitch than the target.

Two test tapes were generated. The pretest, as usual, contained pairs of isolated noises, arranged in six blocks of 18, with ISIs of 750 msec within pairs, of 3 sec between pairs, and of 6 sec between blocks. N3 and N5 were paired with each other and with all other noises, just as in Experiments 3 and 4. In the main test, each of the 80 words appeared five times, once with each probe noise. Half the words (the odd items in Table 8) appeared with the probe noise preceding the sentence, and the others (the even items) appeared with the probe noise following. They were arranged into five blocks of 80 according to a Latin square design, such that each word appeared exactly once in each block, and each probe condition occurred eight times in each block. The sequence was randomized within blocks, with ISIs of 500 msec within pairs (i.e., between the probe noise and the sentence onset or offset), 3 sec between pairs, and longer intervals between blocks.

Procedure. For the pretest, the responses for the first block (18 trials) were filled in on the answer sheet, as in previous experiments. The five blocks of the main test were presented in different orders for different subjects by varying the starting block. After the main test, the first block was presented once again, and the subjects' task was to identify the initial consonant of each word as either /s/ or /ʃ/. Although /s/ and /ʃ/ were suggested as responses, subjects were encouraged to write down any other consonants they might hear. Also, it was emphasized that subjects should write down what they heard, even if that resulted in a nonword.

Results

The results of the pretest are shown in Table 9. The noises were easier to discriminate than those in the previous experiments, or else this group of subjects was just more accurate.

The results of the main test are shown in Table 10. The 80 words were divided into eight groups of 10, according to whether they originally began with /s/ or /ʃ/ (the

Table 9
Average Response Percentages for Isolated Noise Pairs in Experiment 5
(Second Noise Lower Than/Equal to/Higher Than First Noise)

First Noise	Second Noise														
	N1			N2			N3			N4			N5		
N1							8	37	55				0	0	100
N2							2	69	29				0	0	100
N3	94	6	0	71	27	2	13	78	9	2	63	35	0	0	100
N4							65	31	4				2	10	88
N5	100	0	0	100	0	0	94	4	2	65	35	0	4	91	5

cues factor), whether they were lexically ambiguous or unambiguous (ambiguity), and whether the probe noise came first or second (order). The division of nonambiguous words into phonotactically constrained and unconstrained ones is not reflected in the table. There were two highly significant effects: the main effect of probe [$F(4,36) = 74.16, p < .0001$] and the interaction of probe with order [$F(4,36) = 22.03, p < .0001$]. The subjects were quite successful in discriminating the probe noises regardless of whether they came first or second, but they were more accurate when they followed the sentence, as in Experiment 2. The only other effect that reached significance was the ambiguity × probe interaction [$F(4,36) = 3.11, p < .03$], apparently owing to a tendency to perceive the target noise as relatively higher in ambiguous stimuli, but only when the probe noise was N4. This was the probe noise that, on the whole, was perceived as most similar to the target noise (N5), which again replicated the tendency to perceive target noises as lower in pitch than probe noises. (The present design certainly encouraged such a bias.) The cues main effect was nonsignificant [$F(1,9) = 0.70$], though the crucial cues × ambiguity interaction was close to significance [$F(1,9) = 4.84, p < .06$]. This interaction was due to one category of words, ambiguous /s/ words, differing from the rest: In these stimuli, the target noise was judged to be relatively higher than in the other stimuli. This is contrary to expectations, of course; the segregation hypothesis predicted that target noises would be perceived as lower in /s/ words, particularly in unambiguous /s/ words. Although target noises

were perceived as somewhat lower in unambiguous /s/ words than in ambiguous /s/ words, there was no tendency to perceive them as lower in /s/ words than in /ʃ/ words. Thus, the results of this experiment, like those of Experiments 3 and 4, offer no support for the segregation hypothesis.

This negative result carries weight only if listeners in fact restored the fricative consonants intended by the original speaker. Subjects' identification responses suggested that this was generally the case, particularly in the lexically unambiguous stimuli. Unambiguous /ʃ/ words were correctly identified on 97% of the trials when a vowel followed the initial consonant, but on only 76% of the trials when /r/ followed; the errors in the latter case were predominantly F. Unambiguous /s/ words were correctly identified on 95% of the trials, regardless of following context. Ambiguous /ʃ/ words were correctly identified on 82% of the trials (errors being nearly always S), and ambiguous /s/ words on 84% (errors being predominantly SH, and sometimes TH). These percentages indicate that phoneme restoration did occur, and that it was guided both by residual acoustic cues and by a lexical bias.

GENERAL DISCUSSION

The present experiments addressed two conflicting hypotheses concerning the processes underlying phoneme restoration. The hypotheses rest on the basic assumption that an abstract phonological representation composed of phonemes can be distinguished from a concrete auditory percept composed of phones. Given this assumption, the top-down completion hypothesis stated that phoneme restoration occurs at the phonological level as a consequence of bottom-up and top-down constraints, but that it is not accompanied by active phone restoration from the acoustic signal. The phone is restored "in the mind's ear," if at all. Consequently, an extraneous noise replacing a fricative noise should be perceived veridically. The segregation hypothesis, on the other hand, stated that the acoustic material occurring in the place of the fricative noise is partitioned into a speech and a nonspeech portion, either by means of schema-guided auditory scene analysis (auditory induction) or by usurpation of the acoustic signal by a specialized speech module (preemptiveness). Consequently, the extraneous noise replacing [s] frication should be perceived as lower in pitch than an identical noise replacing [ʃ] frication or a stop closure.

Table 10
Average Scores (Percent Judgments of "Probe Higher" Minus Percent Judgments of "Probe Lower") in the Main Test of Experiment 5 (Target Noise was N5)

Words	N1	N2	N3	N4	N5
	Probe First				
Ambiguous /ʃ/	−60	−47	−22	−14	36
Ambiguous /s/	−58	−52	−38	−23	40
Unambiguous /ʃ/	−50	−48	−27	−3	.47
Unambiguous /s/	−64	−52	−32	−5	37
	Probe Second				
Ambiguous /ʃ/	−89	−69	−24	24	89
Ambiguous /s/	−95	−76	−27	4	76
Unambiguous /ʃ/	−90	−76	−26	20	87
Unambiguous /s/	−84	−58	−30	38	76

A probe noise matching task was employed to test these predictions.

Experiments 1 and 2 compared the perception of target noises that replaced either [s] frication or a stop consonant closure silence. The results seemed to favor the segregation hypothesis: Noises replacing [s] were judged to be lower in pitch than noises replacing /t/ silence, relative to an external probe noise. However, the effect was much smaller than it would have been if full restoration of the [s] noise had occurred. Experiments 3 and 4, despite the use of rather similar stimulus materials, did not replicate the effect. One difference was that the /t/ release burst was removed from the control stimuli, in which the extraneous noise replaced the closure silence of /t/. Another difference was that the subjects were not as certain about the initial consonant as in the previous experiments, because of the inclusion of nonword stimuli such as "teminar" and "selephone." Both these changes relaxed the tight constraints on phoneme identity and brought the phone restoration paradigm somewhat closer to the standard phoneme restoration situation, where no residual cues are present in the signal and a large variety of stimuli are employed. Experiment 5 indeed used a large sample of words and compared the perception of target noises replacing /s/ or /ʃ/. As in Experiments 3 and 4, no segregation effect was obtained.

The apparent segregation effect in Experiments 1 and 2 could have been due to an undesired effect of acoustic context in the control stimuli: The /t/ release burst, a brief noise strong in high spectral frequencies, may have been integrated with the target noise into a single noise percept, which naturally sounded higher in pitch. Alternatively, the conjunction of residual signal cues and lexical certainty may indeed have enabled listeners to carry out a segregative process, though it was not very effective. If the process occurs only under such highly constrained conditions, which are not representative of the typical phoneme restoration experiment, its significance would seem to be limited. Yet, it might provide an existence proof of schema-guided influences on primitive auditory scene analysis (Bregman, 1990), or even of preemptiveness of a speech module (Liberman & Mattingly, 1989). This author's inclination is to consider the segregation effect in Experiments 1 and 2 a stimulus artifact.

In conditions that were still quite constrained in comparison with those in other phoneme restoration experiments and that did lead to restoration of the correct fricative consonant on most trials (Experiments 3–5), there was no trace of any segregative or preemptive processes. This negative result supports the top–down completion hypothesis. According to this hypothesis, the target noise aided phoneme restoration by masking the absence (or the presence, as the case may be) of a fricative noise, but it was not partitioned into the missing phone and an acoustic residue. Listeners appeared to hear the full target noise as an extraneous sound.

The great unknown in phoneme restoration experiments is what the restored speech really sounds like. The problem is that subjects' pronouncements about what they are "hearing" cannot be taken at face value; normally, they will be referring to a linguistic abstraction, not to a concrete auditory percept. The auditory percepts accompanying some phonemes (such as stop consonants) are difficult to describe and may be inaccessible to consciousness. For this reason, the present study employed fricatives, whose auditory correlates are relatively accessible. Repp (1981b) demonstrated that, in a fricative discrimination task, naive listeners seem to make judgments on the basis of their phonemic percepts; after some training, however, they are able to direct their attention to the auditory percepts of the fricative noises and judge their timbre quite accurately (presumably without losing the phonemic percept). Fricative consonants, therefore, can be perceived at two levels, one abstract and synthetic, the other concrete and analytic. It made sense, therefore, to ask whether the concrete percept is restored along with the abstract one in the phone restoration paradigm.

If the concrete fricative noise were restored in a top-down fashion, as a consequence of phoneme restoration, one might expect the illusory restored phone to have the prototypical auditory quality associated with the restored phoneme. Indeed, Warren and his collaborators have always claimed that a restored phoneme is auditorily indistinguishable from the original. Informal observations during the present studies suggested that the restored /s/ was subject to quality variations that depended on the acoustic information provided in its place. Obviously, these quality variations are related to the fact that the target phoneme was not always identified correctly. The impression of poor quality may have derived from either a lack of confidence in making the phonemic decision or an awareness of the atypical auditory quality of the noise substituting for the fricative phone (or of its insufficiency as a potential masker of a real [s] noise, which amounts to the same thing). It is noteworthy that, in earlier phoneme restoration studies, the subjects were never asked to identify the restored phoneme or to comment directly on the quality of its sound or articulation. Instead, they were required either to localize the extraneous noise with respect to the speech (Warren's paradigm) or to discriminate "added" from "replaced" items (Samuel's paradigm). It is likely that subjects' above-chance discrimination performance in Samuel's experiments was due to their ability to distinguish quality differences in the critical phoneme.

One reason for the presence of these quality variations may lie in the intensity of the replacement sounds used.[6] In Warren's experiments, these sounds were always much louder than the speech sounds they replaced. Warren (1984) stressed that the replacement sound must be an effective masker of the speech sound for complete restoration to occur. If the extraneous sound masks the speech completely, then "replaced" and "added" conditions are, obviously, indistinguishable. Samuel's discrimination paradigm, however, required that "replaced" and "added" versions of a stimulus be discriminable to some extent,

so that variations in the "strength" of phonemic restoration (as measured by d') as a function of other factors could be investigated. Therefore, Samuel used extraneous sounds that were similar to the critical speech segment in amplitude, and Samuel and Ressler (1986) introduced the use of signal-correlated noise, which retains the exact speech amplitude envelope. These replacement sounds did not mask the speech completely: In speech, spectral energy is unevenly distributed, whereas in signal-correlated noise, the same total energy is spread over the whole frequency range. The peaks of the speech spectrum are therefore not masked by the noise.

The present study followed Samuel's methods, for a similar reason: If a very loud noise had been used to replace [s], as in Warren's experiments, any effect of perceptual segregation on its perceived timbre surely would have been too small to show up in the noise matching task. By using replacement noises whose intensity was comparable to that of the [s] noise to be restored, chances for observing a segregation effect were improved—or so it was thought. This expectation would hold only if segregation were guided by top-down expectancies and if a prototypical [s] spectrum were subtracted from the replacement noise. It appears, however, that this is not what happened; rather, the spectral quality of the perceived phone (if any) seemed to resemble that of the replacement noise. (No formal judgments of the sound quality of the fricative phone were obtained.)

It is possible that the present findings are specific to the stimuli used here. Not only were the replacement noises of the same intensity as the original fricative noise (as in Samuel's studies), but also they offered absolutely no bottom-up clues to spectral segregation, even when the original noise was part of the mixture, because of the unitary nature of random noise waveforms. It appears that auditory scene analysis cannot be driven entirely by top-down expectations, but the possibility remains that segregation processes triggered by the input are enhanced by top-down processes. It is relevant in that connection to note that the detectability of masked speech is not increased by knowledge of the speech to be detected, conveyed through either reading (Frost, Repp, & Katz, 1988) or lipreading (Repp, Frost, & Zsiga, in press). Also, lipreading information does not enhance phoneme restoration in Samuel's paradigm; in fact, it seems to interfere (Trout & Poser, 1990). These negative findings suggest that Bregman (1990) is right in surmising that the processes of primitive auditory scene analysis are autonomous.

The findings of Whalen and Liberman (1987) and Bentin and Mann (1990) provide a challenge (see Bregman, 1990, 1991, for discussion). They showed that a frequency transition embedded in speech contributes to speech intelligibility at intensities far below those at which the transition becomes detectable as a whistle or a chirp. It seems as if the speech system appropriated most of the transition's energy, leaving only a residue for the auditory system. However, the finding may also be understood as showing that speech is perceived holistically (cf. Warren,

1983), whereas analytic perception of a speech component as "standing out" requires bottom-up cues to source segregation, such as a raised intensity. The fact that a transition segregated in this manner continues to contribute to speech perception can be understood as an instance of self-masking; that is, any auditory event may contain a weaker copy of itself. This is very similar to the situation in fricative perception, where the unmodified noise may be heard as a separate source with primitive auditory qualities (Repp, 1981b), without any apparent effect on perception of the fricative phoneme.

Two possible criticisms of the present research need to be addressed. One is that the subjects were simply unable to detect the target noise in the speech and responded on the basis of the probe noise only. Although this seems possible with regard to Experiments 3 (which used only following probes) and 5 (which used a single target noise), other experiments showed that listeners were sensitive to differences among target noises varying in spectrum. More importantly, subjects never complained that they were unable to perform the task because they heard no noise in the sentence. There is little doubt that they did perform a noise comparison task, even though the timbre of the target noise was not easy to judge. The other criticism is that the stimuli were not conducive to phoneme restoration: The target noise served as a poor fricative noise, and subjects perceived whatever phoneme was most compatible with the acoustic evidence. This may describe the situation quite accurately. However, there is a continuum of possible scenarios between two extremes: on the one hand, Warren's classic demonstration in which th. extraneous noise is not speechlike and is more intense than the speech signal it replaces, and, on the other hand, the original, unmodified speech signal (in which the extraneous noise is identical with the phone it replaces). It is not clear where phoneme restoration ends and ordinary phoneme perception begins. Restoration essentially refers to the apparent intactness of the speech, and more narrowly to the perception of the phoneme intended by the speaker. It is difficult to draw a line that excludes phoneme perception based on the original signal as a legitimate instance of "restoration." Therefore, any objection to the present experiments (and to Samuel's similar paradigm) on the grounds that they did not lead to phoneme restoration is void. Moreover, the present Experiments 3–5, in contrast with most earlier experiments, actually provided evidence that the restored phonemes corresponded to those intended by the speaker, most of the time anyway.

Applying the notion of restoration to the extreme situation of self-replacement (i.e., an unmodified speech signal) is instructive with regard to evaluating the hypothesis of *phone* restoration, which was under test here. It would be bizarre to suggest that the original [s] noise needs to be segregated into two identical copies to support both phoneme and timbre perception. Clearly, it is just a matter of attending to one or the other of its two aspects: the concrete, auditory percept or the abstract, phonological one (for which it constitutes only one of a number of in-

formation sources). If an original [s] noise is replaced with a noise that has different spectral properties (and this includes any noise obtained by adding some noise to an original [s] noise, for two added noises are perfectly fused and form a single sound), listeners will presumably adopt the same two perspectives, but with different degrees of emphasis. The less similar the replacement noise is to an [s] noise (or any other plausible fricative noise) in spectrum and intensity, the less it will contribute to phoneme perception, and the more readily it will be perceived as an extraneous sound. The present experiments apparently represented situations in which the replacement noise was sufficiently similar to [s] to contribute to phoneme identification, but also sufficiently dissimilar to be perceived as extraneous without much attentional effort. In Warren's experiments, the intense replacement noise presumably did not contribute to phoneme identification, which was entirely top-down, and was perceived as strikingly separate from the speech.

Phonological perception does not require awareness of specific auditory components. Some phones can be attended to as sound objects if they are in the acoustic input; if they are masked, they become indefinite. Restored phonemes are no less real if they have no accessible auditory substrate; after all, stop consonants are perceived this way all the time. Fricatives are special because their principal auditory component is relatively accessible as an independent percept. It seems, however, that fricatives can be perceived just as well if their characteristic noise is inaccessible or truly absent, as in the sine-wave speech stimuli of Remez and Rubin (1990), as long as other bottom-up and top-down constraints are sufficient for phoneme perception. Phone restoration may not occur because perception of auditory segments as such is not a prerequisite for phoneme perception. The latter point, of course, has long been made by Liberman and his associates (see, e.g., Liberman & Mattingly, 1989) and, in different ways, by Warren (1983) and Bregman (1990) as well. Nevertheless, confusion is perpetuated by frequent references to phonemes as "speech sounds" in the literature, as if they needed to be heard. The belief that restored phonemes are "heard" is closely connected to this usage. Acceptance of phonemes as abstract linguistic units without specific auditory properties may go a long way toward clarifying the intriguing phenomenon of phoneme restoration.

REFERENCES

BENTIN, S., & MANN, V. A. (1983). Selective effects of masking on speech and nonspeech in the duplex perception paradigm. *Haskins Laboratories Status Report on Speech Research*, SR-76, 65-85.

BENTIN, S., & MANN, V. A. (1990). Masking and stimulus intensity effects on duplex perception: A confirmation of the dissociation between speech and nonspeech modes. *Journal of the Acoustical Society of America*, 88, 64-74.

BREGMAN, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

BREGMAN, A. S. (1991). The compositional process in cognition with applications to speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 197-223). Hillsdale, NJ: Erlbaum.

BREGMAN, A. S., & PINKER, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32, 19-31.

ELMAN, J. L., & MCCLELLAND, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, 27, 143-165.

FOSS, D. J., & BLANK, M. A. (1980). Identifying the speech codes. *Cognitive Psychology*, 12, 1-31.

FROST, R., REPP, B. H., & KATZ, L. (1988). Can speech perception be influenced by simultaneous presentation of print? *Journal of Memory & Language*, 27, 741-755.

GLAVE, R. D. (1973). *Untersuchungen zur Tonhöhenwahrnehmung stochastischer Schallsignale*. Hamburg: Helmut Buske.

HESSE, H.-P. (1982). The judgment of musical intervals. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 217-226). New York: Plenum.

LIBERMAN, A. M., & MATTINGLY, I. G. (1989). A specialization for speech perception. *Science*, 243, 489-494.

MANN, V. A., & LIBERMAN, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, 14, 211-235.

MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 28, 213-228.

MANN, V. A., & REPP, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548-558.

MANN, V. [A.], & SOLI, S. D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics*, 49, 399-411.

MARSLEN-WILSON, W. (1987). Functional parallelism in spoken word-recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 71-102). Cambridge, MA: MIT Press.

MATTINGLY, I. G., & LIBERMAN, A. M. (1988). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 775-793). New York: Wiley.

MATTINGLY, I. G., & LIBERMAN, A. M. (1990). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Signal and sense: Local and global order in perceptual maps* (pp. 501-520). New York: Wiley.

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

NUSBAUM, H. C., SCHWAB, E. C., & SAWUSCH, J. R. (1983). The role of "chirp" identification in duplex perception. *Perception & Psychophysics*, 33, 323-332.

REMEZ, R. E., & RUBIN, P. E. (1990). On the perception of speech from time-varying acoustic information: Contributions of amplitude variation. *Perception & Psychophysics*, 48, 313-325.

REPP, B. H. (1981a). On levels of description in speech research. *Journal of the Acoustical Society of America*, 69, 1462-1464.

REPP, B. H. (1981b). Two strategies in fricative discrimination. *Perception & Psychophysics*, 30, 217-227.

REPP, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 243-335). New York: Academic Press.

REPP, B. H. (1989). Phone restoration. In T. Szende (Ed.), *Proceedings of the Speech Research '89 International Conference* (pp. 238-241). Budapest: Linguistics Institute of the Hungarian Academy of Sciences.

REPP, B. H., FROST, R., & ZSIGA, E. (in press). Lexical mediation between sight and sound in speechreading. *Quarterly Journal of Experimental Psychology*.

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

SAMUEL, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 1124-1131.

SAMUEL, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory & Language*, 26, 36-56.

SAMUEL, A. G. (1991). A further examination of attentional effects in the phonemic restoration illusion. *Quarterly Journal of Experimental Psychology*, 43A, 679-699.

SAMUEL, A. G., & RESSLER, W. H. (1986). Attention within auditory word perception: Insights from the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception & Performance*, 12, 70-79.

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

SOLI, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 976-984.

TROUT, J. D., & POSER, W. J. (1990). Auditory and visual influences on phonemic restoration. *Language & Speech*, 33, 121-136.

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

WARREN, R. M. (1976). Auditory illusions and perceptual processes. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 389-417). New York: Academic Press.

WARREN, R. M. (1983). Multiple meanings of "phoneme" (articulatory, acoustic, perceptual, graphemic) and their confusions. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 285-311). New York: Academic Press.

WARREN, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, 96, 371-383.

WARREN, R. M., & OBUSEK, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, 9(3B), 358-362.

WARREN, R. M., OBUSEK, C. J., & ACKROFF, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, 176, 1149-1151.

WARREN, R. M., & SHERMAN, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, 16, 150-156.

WHALEN, D. H., & LIBERMAN, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, 237, 169-171.

NOTES

1. Warren (personal communication) does not view the two hypotheses contrasted here as being mutually exclusive. He seems to reject the notion of an abstract level of phonological representation, at least as a possible site of phonemic restoration; instead, he regards syllables and words as holistic *auditory* patterns (cf. Warren, 1983). The distinction between concrete auditory and abstract schematic representations—mistaken though it may be—is fundamental to the present argument, however.

2. The preemptiveness hypothesis was developed in connection with a task that required spectral completion of an isolated speech signal, whereas phonemic restoration requires temporal completion and is contextually guided. Possibly, therefore, the preemptiveness hypothesis, as conceived by Liberman and his colleagues, does not apply here.

3. Due to characteristics of the original [s] noise, the "happiness" noises were lower in overall intensity than the "seminar" noises and had a somewhat irregular amplitude envelope. These amplitude differences cannot be seen in Figure 1, which shows amplitude-normalized plots of the spectra. (It would have been desirable to include the original [s] spectra in the figure, but they could not be located.)

4. In addition to these main findings, the ANOVA revealed a number of significant interactions that are of little theoretical interest and whose discussion is omitted to ease the burden on the reader.

5. One subject's comment suggests a possible explanation: "Selephone" is rather similar to the real word "cellophane" and may have activated that lexical entry.

6. I am grateful to Richard Warren (personal communication) for drawing my attention to this important point.