

Phonemic restorations based on subsequent context*

RICHARD M. WARREN and GARY L. SHERMAN

University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201

Earlier experiments have shown that when one or more speech sounds in a sentence are replaced by a noise meeting certain criteria, the listener mislocalizes the extraneous sound and believes he hears the missing phoneme(s) clearly. The present study confirms and extends these earlier reports of phonemic restorations under a variety of novel conditions. All stimuli had some of the context necessary for the appropriate phonemic restoration following the missing sound, and all sentences had the missing phoneme deliberately mispronounced before electronic deletion (so that the neighboring phonemes could not provide acoustic cues to aid phonemic restorations). The results are interpreted in terms of mechanisms normally aiding veridical perception of speech and nonspeech sounds.

This study is concerned with processes which normally aid auditory perception under noisy conditions. If an item in an auditory sequence is replaced by a louder sound having an appropriate spectrum, both the missing sound and the extraneous sound may be heard clearly. This illusory perception of missing sounds has been called "auditory induction" (Warren, Obusek, & Ackroff, 1972). The product of auditory induction seems quite real, and a restored sound cannot be distinguished perceptually from sounds actually present.

A special type of auditory induction called "phonemic restorations" is evoked when a portion of speech is replaced by noise. In the only detailed experiments published to date, the missing speech sounds within a sentence could be identified by the preceding context (Warren, 1970; Warren & Obusek, 1971). When the "s" or the "gis" in the word "legislatures" was replaced by a loud extraneous sound (such as a cough), the listener thought the sentence was intact and mislocalized the noise, believing that it coexisted with some other portion of the sentence (usually several phonemes earlier than the actual position). When the same sentence was used with all phonemes present and a short click in the middle of the "s," this extraneous sound was localized considerably later in the sentence by those hearing the sentence for the first time. It was hypothesized that both direction and extent of errors depended in part upon the time required for perceptual synthesis in phonemic restorations.

Some preliminary experiments in our laboratory have indicated that listeners could store ambiguous information and perceptually restore phonemes on the basis of context following the deleted portion. In the present study, information necessary for the proper phonemic restoration followed the missing sound in each case. In order to ensure that acoustic cues indicating the

identity of the missing phoneme were not supplied by neighboring speech sounds, the target phoneme was deliberately mispronounced on the original recording before deletion. The removal and addition of sounds were electronically controlled with millisecond precision.

METHOD

Subjects

The 60 Ss were students from the introductory psychology course at the University of Wisconsin-Milwaukee. They were either paid \$1 or received credit toward their course grade, depending upon their wishes and the rules of their particular instructor. Each served in only one of the three groups of 20 Ss. They all had no known history of auditory impairment and no prior experience with experiments dealing with localization of events in sentences.

Stimuli

In previous studies of phonemic restorations, stimuli were produced by cutting and splicing tape-recorded sentences. The procedure employed in this study was novel, and much more flexible. The "cuts" were controlled electronically and were reversible. They could be introduced with reproducibility of $\pm 1/2$ msec relative to positions within a recorded sentence.

The seven stimulus sentences listed in Table 1, each containing one inappropriate phoneme, were produced clearly at a normal rate in an audiometric room by a male talker speaking general American English. The microphone of a General Radio Model 1551C sound-level meter was about 12 in. from his mouth. The meter scale reading permitted monitoring of level by the talker, and the output of the sound-level meter was used for recording outside the audiometric room on an Ampex AG-440B at 15 in./sec. A 2,000-Hz square wave had been recorded previously on Track 2. The sections of tape containing each sentence were cut out and spliced together with blank leader tape separating the sentences. On playback of Channels 1 and 2 on an Ampex AG-500 recorder, acoustic events within the sentence could be located in 0.5-msec steps by counting the number of pulses from the onset of the sentence. The counting was accomplished with a Hunter Model 1516 multichannel counter. When a preset number of counts was reached, a switching signal was sent to a Grason-Stadler Model 829 E electronic switch, causing the output to be changed from Channel A to Channel B. The voice track of the recording went to Channel A (there was no input to Channel B at this point). The output from the electronic switch went to a McIntosh Model MA 5100 preamplifier/amplifier, and

*This study was supported in part by a University of Wisconsin-Milwaukee Graduate School grant to the first author.

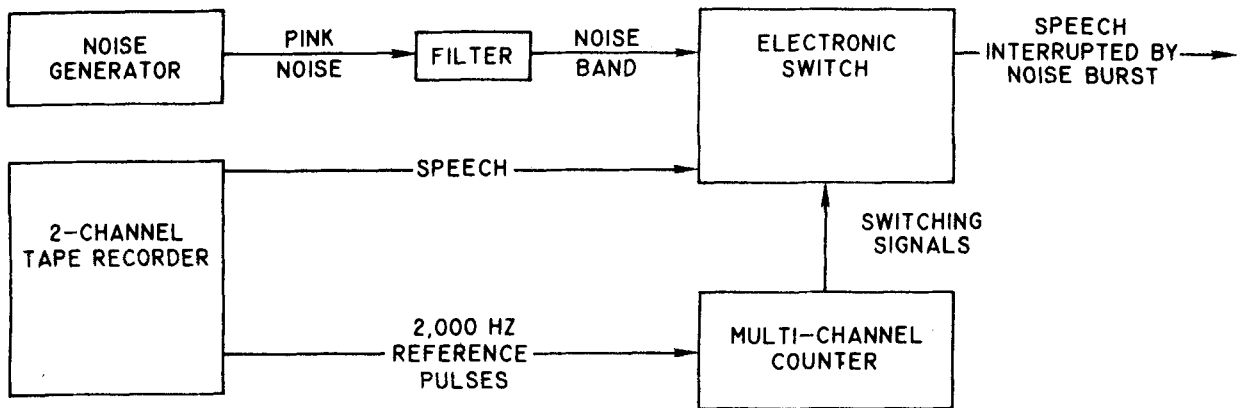


Fig. 1. Apparatus used for "electronic splicing" of interruptions into sentences.

thence to a matched pair of TDH 49 headphones. The E could hear the sentence clearly up to the shut-off point (which was under his control). As had been noted earlier (Warren, 1971), if a recorded sentence is abruptly shut off before completion, the phoneme marking termination is readily identified. With the sentences used in this study, it was quite easy to perceive whether termination occurred at the beginning, middle, or end of the target phoneme. The counter was adjusted by the E to terminate the sentence at the end of the phoneme just preceding the mispronounced speech sound. The number of counts required to terminate the sentence at this position was then determined independently by the other E for each sentence. These two values agreed within 10 pulses (5 msec) for each sentence. Next, the counter was set for a different control mode, so that the beginning of the sentence was not heard, and the electronic switch was triggered not to turn on the sentence until a preset number of pulses had been counted. The count required to initiate the sentence at the beginning of the phoneme which immediately followed the mispronounced sound was determined independently by each E. Identification of the initial sound of a sentence fragment was almost as easy as identification of the terminal sound had been, and again Es agreed within 5 msec.

Using the termination and onset values for each sentence, it was possible to set the multichannel counter so that all of the sentence was heard except the misstated phoneme. In order to minimize the immediate transitional cues linking adjacent speech sounds, the deleted portion of the sentence was expanded 20 to 40 msec in each direction, so that interruption began and ended well within the phonemes surrounding the target, and the duration of the deleted portion (neglecting rise/decay switching times) was about 150 msec. In preparing the stimuli for Experiment 3, in which a silent gap replaced the mispronounced phoneme (and portions of the adjacent phonemes), the multichannel counter located the desired positions in the sentences and caused the electronic switch to delete the preselected segment (rise/decay setting was 5 msec to minimize switching transients). The total durations for deletions for each sentence are given in Table 1 (calculated from halfway points of signal rise and decay). The sentences having silent gaps were recorded on another Ampex Model AG-500 recorder and used as stimuli in Experiment 3.

In Experiments 1 and 2, a noise was introduced to replace the deleted portion of the sentence. This "electronic splicing" was accomplished as indicated in Fig. 1. The intact original recording of the stimulus sentence was delivered to one of the two input channels of the electronic switch. Pink noise (equal energy per octave) was produced by a General Radio Model 1382 generator, sent through a Rockland filter Model 1042F adjusted to pass a band from 100 through 4,000 Hz (cutoff slope at limits equal to 48 dB per octave), and thence to the other channel of the

switch. The multichannel counter triggered the electronic switch (rise/decay setting at 5 msec) to delete the same portions of the sentences as for the stimuli used in Experiment 3, except that a 100- to 4,000-Hz burst of pink noise replaced the deleted portion of the sentence. The sentences were rerecorded along with the noise burst. The noise burst level was set 10 dB above the peak level of the intact portion for each sentence.

Table 1 provides an estimate of the level of the noise burst relative to the restored portion of the sentence. In order to make this estimate, each sentence was rerecorded by the same talker with the correct phoneme rather than the incorrect phoneme used for preparing the experimental stimuli. A 2,000-Hz square wave was recorded on the second channel, and the equipment used in preparation of the experimental stimuli and shown in Fig. 1 was used to permit passage only of that portion of each reference sentence corresponding to the deleted portions of the stimuli (the segments perceptually synthesized during phonemic restorations). The intensities of these segments were measured using a Bruel and Kjaer Model 2204 sound-level meter in the impulse (peak hold) mode. Next, the entire reference sentence was played, and the peaks of intensity measured with the sound-level meter in the normal real-time response mode. Since the noise level was set at 10 dB above the peak level of the intact portion of each experimental sentence, adding 10 dB to the measured sentence peak of the reference sentence and subtracting the measured level of the phoneme corresponding to phonemic restoration gave an approximate value for the difference between the noise and the phonemic restoration levels in the experimental sentences. These values are given in Table 1 in the column headed: Noise-PhR Level (decibels). It should be emphasized that the sentences pronounced correctly were used only for this calculation; the sentences pronounced with the incorrect phonemes were used in preparation of all stimuli.

Procedure

Each S was tested separately in an audiometric room. In all three experiments, Ss received the sentences in Table 1 in the order listed. The stimuli were heard diotically through matched headphones at a voice-peak level of 80 dB. In Experiment 1, Ss read the following instructions: "You will hear a tape recording of a sentence spoken clearly. A noise will occur at some time during the sentence. You are to indicate where you think the noise occurred by placing a circle around the *exact* place in a typewritten statement of the sentence. You are to indicate also whether the noise completely replaced any sound(s) which you circle." After the S indicated that the instructions were understood, he heard the first sentence containing the 154-msec noise burst at 90 dB. Immediately after hearing the stimulus, the S was given a typewritten statement of the complete meaningful sentence with single spaces separating letters in words and

Table 1
Stimuli Used to Test for Phonemic Restorations With the Three Experimental Groups

Sentence	Sentence Duration (msec)	Deletion Duration (msec)	Noise-PhR Level (dB)
1 It is common for people to jump on the [s/b] and wagon when a political movement becomes popular.	3930	154	+18
2 George waited for the deli[b/v] ery of his new color TV.	2690	155	+16
3 George decided that commu[t/n] icating with his children was very difficult.	3200	154	+15
4 George waited for the deli[v/b] eration of the indecisive jury.	2860	155	+18
5 It was the circum[f/st] ances surrounding the crime which caused so much concern.	3220	155	+23
6 George remembered the Dec[ed/em] ber of '68 that produced the giant snowstorm.	3430	156	+14
7 Many felt that the senator went [s/b] eyond all bounds in his blast at the administration.	3870	155	+24

Note—The deleted portions of the sentences are indicated by brackets; within the brackets, the actual stimuli before deletion are indicated first, followed by the sounds which correspond to phonemic restorations.

double spaces between words. He circled the position at which he thought the noise occurred and then noted whether he thought the noise completely replaced the circled sound(s). The answer sheet was removed, the same sentence presented for a second time, and a fresh answer sheet presented. This procedure was repeated once more, so that a total of three responses was obtained from each S before presenting the next sentence. This process was continued until three responses were obtained for each of the seven sentences.

In Experiment 2, Ss heard the same stimuli but were informed that a part of the sentence was missing. The instructions stated,

"You will hear a tape recording of a sentence spoken clearly. A portion of the recorded sentence has been cut out and replaced by a recorded noise. You are to indicate the missing sound(s) by placing a circle around the exact place in a typewritten statement of the sentence." The answer sheets were the same as in Experiment 1, except that Ss were not asked whether the noise replaced the circled speech sound(s).

In Experiment 3, Ss heard the sentences with silent gaps replacing the missing sounds rather than noise. The instructions for this experiment were, "You will hear a tape recording of a sentence spoken clearly. A silent gap will occur at some time

Table 2
Error Magnitudes* for Temporal Localization by Groups of 20 Listeners, Each Making Three Successive Judgments

Sentence	Experiment 1** Judgment			Experiment 2**, †† Judgment			Experiment 3 † Judgment		
	1	2	3	1	2	3	1	2	3
1 Phonemes (Mdn)	6.25	2.00	1.42	1.75	2.38	2.35	1.50	0.62	0.58
Msec (Mdn)	295	121	102	120	104	101	108	56	56
Msec (Mn)	340.2	128.7	115.3	336.9	180.2	128.4	116.8	69.8	56.6
SE of Mean	66.2	18.7	17.0	92.8	43.0	31.1	23.8	14.4	11.6
2 Phonemes (Mdn)	4.25	2.62	1.12	4.50	2.11	1.62	3.25	1.25	1.55
Msec (Mdn)	225	156	100	228	131	116	183	100	105
Msec (Mn)	250.9	186.2	114.9	262.2	156.2	134.0	204.6	113.0	107.7
SE of Mean	38.5	32.4	19.0	37.3	16.5	16.4	32.5	21.7	17.4
3 Phonemes (Mdn)	3.00	2.05	1.75	2.08	1.42	1.42	1.50	1.12	1.04
Msec (Mdn)	182	134	125	134	120	115	115	95	90
Msec (Mn)	184.0	133.8	128.8	159.5	127.7	120.1	122.9	91.4	86.9
SE of Mean	18.6	13.2	11.3	18.1	21.7	20.1	16.0	11.3	10.8
4 Phonemes (Mdn)	1.50	1.00	1.00	1.92	1.33	0.83	1.12	1.12	0.54
Msec (Mdn)	126	102	80	141	116	81	102	98	42
Msec (Mn)	151.1	116.2	104.6	164.7	114.9	93.6	89.8	88.8	64.8
SE of Mean	30.5	25.7	22.2	25.8	12.7	16.5	12.7	13.2	13.8
5 Phonemes (Mdn)	2.00	1.75	1.62	1.75	0.65	0.55	0.92	0.21	0.18
Msec (Mdn)	140	131	126	136	54	63	94	22	18
Msec (Mn)	167.9	131.1	117.4	153.7	88.7	69.7	94.2	46.9	31.5
SE of Mean	32.9	20.2	18.1	29.5	20.7	14.5	20.9	14.5	12.0
6 Phonemes (Mdn)	0.75	0.62	0.57	1.12	0.25	0.33	0.38	0.21	0.19
Msec (Mdn)	93	89	70	144	34	64	64	25	23
Msec (Mn)	115.9	97.9	92.9	152.1	53.9	51.8	75.5	62.0	47.2
SE of Mean	22.1	19.0	19.7	17.2	15.5	12.0	18.9	19.4	17.5
7 Phonemes (Mdn)	3.58	0.92	1.00	3.00	1.38	1.12	1.34	0.75	0.85
Msec (Mdn)	183	80	88	170	122	77	79	70	73
Msec (Mn)	293.6	149.0	116.2	247.5	128.1	110.3	214.2	62.7	67.9
SE of Mean	51.1	32.0	25.3	48.7	19.8	21.6	57.0	13.0	14.5
Grand Mean	214.8	134.7	112.9	210.9	121.4	101.1	131.1	76.4	66.1

*Direction of errors not considered; early and late localizations both considered positive.

**Localization judgments involved noise bursts replacing phonemes.

†Localization judgments involved silent gaps replacing phonemes.

††Listeners told that noise completely replaced one or more speech sounds.

Table 3
Temporal Direction Errors* by Separate Groups of 20 Listeners, Each Making Three Successive Judgments

Sentence	Experiment 1** Judgment			Experiment 2**, †† Judgment			Experiment 3 † Judgment		
	1	2	3	1	2	3	1	2	3
1 Phonemes (Mdn)	-5.50	-1.75	-0.92	-2.25	-1.75	-1.00	-0.25	-0.11	-0.12
Msec (Mdn)	-175	-91	-80	-98	-91	-88	-22	-10	-10
Msec (Mn)	-181.0	-65.7	-45.6	-195.5	-79.0	-59.7	-3.1	-14.7	-22.8
SE of Mean	92.2	33.8	29.6	112.3	56.2	40.7	35.9	21.3	16.6
2 Phonemes (Mdn)	-4.25	-2.50	-0.92	-4.50	-2.00	-1.25	-3.25	-1.08	-1.25
Msec (Mdn)	-227	-142	-87	-232	-115	-100	-183	-97	-100
Msec (Mn)	-250.9	-136.4	-90.2	-262.2	-115.6	-87.5	-204.6	-83.7	-77.0
SE of Mean	38.5	43.6	25.5	37.3	29.5	28.5	32.5	27.9	23.9
3 Phonemes (Mdn)	-3.00	-2.05	-1.58	-2.08	-1.00	-0.75	-1.50	-1.12	-1.00
Msec (Mdn)	-182	-133	-114	-134	-92	-69	-111	-97	-92
Msec (Mn)	-167.9	-128.5	-93.0	-145.7	-45.6	-13.2	-111.7	-86.0	-76.1
SE of Mean	25.4	15.8	23.4	23.4	34.9	34.0	19.8	13.3	14.4
4 Phonemes (Mdn)	-1.75	-0.42	-0.75	-1.75	-1.12	-0.45	-1.08	-1.12	-0.42
Msec (Mdn)	-129	-40	-71	-129	-100	-43	-99	-100	-40
Msec (Mn)	-127.4	-53.4	-66.3	-118.3	-66.6	-20.7	-62.9	-83.4	-48.7
SE of Mean	35.9	34.9	29.0	36.6	24.9	26.6	19.4	14.9	16.9
5 Phonemes (Mdn)	-1.25	-1.00	-0.25	-1.25	-0.50	-0.18	-0.75	-0.12	-0.11
Msec (Mdn)	-114	-105	-26	-114	-52	-19	-79	-13	-12
Msec (Mn)	-105.1	-63.4	-26.4	-119.5	-62.1	-35.9	-77.7	-35.9	-31.5
SE of Mean	44.5	33.2	31.8	36.9	25.3	19.9	24.2	16.0	12.0
6 Phonemes (Mdn)	-0.45	-0.33	+0.25	-1.12	-0.15	-0.14	-0.25	-0.12	-0.10
Msec (Mdn)	-61	-44	+29	-146	-20	-19	-34	-16	-14
Msec (Mn)	-67.2	-41.5	+12.6	-152.1	-18.8	-22.4	-69.6	-56.1	-41.4
SE of Mean	31.0	27.8	28.9	17.2	19.4	16.1	20.0	20.7	18.3
7 Phonemes (Mdn)	-2.75	-0.25	-0.12	-2.50	-0.75	-0.75	0.00	-0.11	-0.18
Msec (Mdn)	-132	-20	-9	-120	-59	-59	0	-9	-14
Msec (Mn)	-125.8	-36.2	-32.9	-80.8	-48.3	-6.5	+73.5	-12.7	-19.3
SE of Mean	84.2	46.2	36.3	72.6	33.7	32.6	73.3	19.2	20.8
Grand Mean	-146.5	-74.9	-48.8	-153.4	-62.3	-35.1	-65.2	-53.2	-45.3

*Early errors considered negative, late errors considered positive.
 **Localization judgments involved noise bursts replacing phonemes.
 †Localization judgments involved silent gaps replacing phonemes.
 ††Listeners told that noise completely replaced one or more speech sounds.

during the sentence. You are to indicate where you think the silent gap occurred by placing a circle around the exact place in a typewritten statement of the sentence. You are to indicate also whether the silent gap completely replaced the sound(s) which you circle." The answer sheets had provisions for circling the portion of the sentence where the gap occurred and also asked the S to indicate whether the gap completely replaced the circled sound(s).

RESULTS

The initial scoring procedure was similar to that employed by Warren and Obusek (1971). In order to measure the errors in the judged position of the extraneous sound or gap, the letter or letters corresponding to individual phonemes on the response sheet were each counted as one position, and the total number of positions separating the circled portion of the printed sentence from the true location was calculated. When more than one position was circled, then the score for each of these was averaged to give a single value for the response. The median error scores in phonemic units are summarized in Tables 2 and 3. In Table 2, for error magnitude, the direction of error (i.e., whether the reported position preceded or followed the actual

position) was ignored, and all deviations were given the same sign. In Table 3, for temporal direction of errors, responses were scored as positive if they followed the actual position and negative if they preceded this position. Hence, it is possible for scores to be high in Table 2 and low in Table 3 if the errors were distributed fairly symmetrically on both sides of the correct position.

While scores expressed in phonemic units permit comparison of results obtained under different conditions with the same stimulus, comparisons across stimuli could be misleading due to the different durations corresponding to the same phonemic scores. Values for localization errors were calculated in milliseconds by obtaining the differences between response position and the true location of the extraneous sound or gap. Timing was accomplished by counting the number of 2,000-Hz pulses on the reference channel accompanying the recording of the stimulus sentence. The numbers of pulses from the onset of the sentence to the beginning and to the end of the target were each counted using the apparatus shown in Fig. 1 and described in the Procedure section. The mean of the limiting values (i.e., the midpoint) was taken as

the temporal location of the target. More specifically, for each response by a S, the number of 0.5-msec pulses corresponding to the beginning and end of the circled phoneme(s) was determined, and the average (midpoint) value was taken as the response position. The response position minus the true target position then divided by two gave the error score in milliseconds. These individual scores were used to calculate the medians, means, and standard errors given in Tables 2 and 3.

Table 2 shows that for each of the statistical measures, the magnitude of absolute errors was generally greatest when the sentences were heard for the first time. Table 3 shows that the extent of the predominantly negative errors in localization (preposing the extraneous sound) was also generally greatest for the first presentation of the sentence. With both of these measures, a marked improvement generally occurred for the second judgment, but once the listeners had heard a sentence twice, presenting it for the third time had little effect upon accuracy.

Looking at the effects of prior stimulation with the same sentence in greater detail, Table 2 shows that a decrease in median error magnitude (in milliseconds) for Judgment 2 occurred for all seven sentences in each of the three experiments. In each experiment, the chance of all seven medians shifting in the same direction is $1/2^6$, so $p = 1/64$. An analysis of individual responses used for construction of Table 2 showed that of those Ss changing judgments for the second presentation, the number decreasing was significantly greater than the number increasing their error magnitude ($p < .05$ by binomial expansion): in Experiment 1 for Sentences 1, 2, 3, and 7; in Experiment 2 for Sentences 2, 5, 6, and 7; and in Experiment 3 for Sentences 1, 2, 5, and 7.

Table 3 shows that the median directional displacement in milliseconds for Judgment 2 was less negative (less preposing of location) than Judgment 1 in Experiment 1 for 7 of 7 sentences; in Experiment 2 for 7 of 7 sentences; and in Experiment 3 for 5 of 7 sentences. This tendency towards less negative displacement was significant for Experiments 1 and 2 ($p < .02$ by binomial expansion). An examination of individual responses used in construction of Table 3 showed that of those Ss reporting changes in localization for the second judgment, the number moving in the direction of later placement was significantly greater than those in the other direction ($p < .05$ by binomial expansion): in Experiment 1 for Sentences 2, 3, and 4; in Experiment 2 for Sentences 2, 3, 4, 5, and 6; and in Experiment 3 for Sentences 2 and 5.

The extent of errors shown in Tables 2 and 3 were similar for Experiments 1 and 2, indicating that the disclosure to Ss that speech sounds were missing provided in Experiment 2 did not influence appreciably their accuracy of judgments. However, the presence of a silent gap in Experiment 3 (rather than an extraneous sound as in Experiments 1 and 2) did enable listeners to locate the interruption in the sentence with greater

accuracy. A sentence-by-sentence comparison in Tables 2 and 3 shows that the median extent of errors for Ss unfamiliar with the sentences (first judgments) was less with Experiment 3 than with Experiments 1 and 2 for each of the seven sentences. The chance of all seven differences between means being in the same direction by chance is $1/64$, so that the differences between experiments are significant for absolute errors and directional errors. The greater accuracy of localization with the gap is in keeping with the findings of Warren and Obusek (1971).

A procedure was devised for dichotomous scoring (accurate or inaccurate) of individual responses in localization of the interruption (noise or silent gap) in the three experiments. Many Ss circled more than one position, considering that the interruption coincided with more than one phoneme. Actually, since portions of the speech sounds adjacent to the deleted phoneme were also removed, such responses were not necessarily completely in error. It was decided to score a response as correct if a S's circle included the missing phoneme and contained no more than a total of three phonemes. An analysis was made of the accuracy of localization using the number of correct localizations in each experiment collapsed over the seven sentences for each of the three judgments. With this analysis, each S could have a score from 0 to 7 for his first, second, and third judgment, respectively. A 3 by 3 split-plot factorial analysis of variance showed a significant effect of experiments [$F(2,57) = 3.69$; $p < .05$] and judgments [$F(2,114) = 21.22$; $p < .01$]. Since the Experiments by Judgments interaction was negligible [$F(4,114) = 0.42$], comparisons of differences between experiments collapsed across judgments and comparisons of judgments collapsed across experiments were carried out using the Newman-Keuls test. The total correct localizations for Experiments 1, 2, and 3 were 118, 108, and 182, respectively, with 182 being significantly greater than 108 and 118 ($p < .01$). Thus, the silent gap enabled listeners to make significantly more correct localizations than did either of the noise conditions, while the two noise conditions did not differ from each other. The total numbers of correct localizations for Judgments 1, 2, and 3 were 99, 150, and 159, respectively, with 99 being significantly less ($p < .01$) than each of the other values. Hence, significant improvement with practice was limited to the first repetition of the stimulus. The data used in the analysis above are shown graphically in Fig. 2.

An analysis of the number of correct identifications of the missing sounds was undertaken. In order to be scored as a correct identification, the S must not only have localized the noise accurately (as defined above and shown in Fig. 2), but must also have stated that the noise replaced the speech sound(s). Experiment 2 was excluded from this analysis since Ss had been informed that a portion of the sentence was missing and replaced with noise. Data were collapsed over the seven sentences

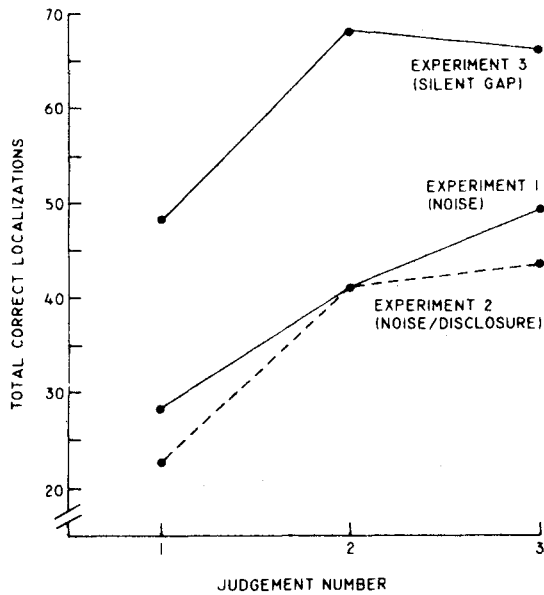


Fig. 2. Total number of correct localizations of the target (noise or silent gap) in each of the three experiments.

for each of the three judgments in Experiments 1 and 3. A split-plot analysis of variance showed no significant effects of judgments nor a significant Experiments by Judgments interaction, but did show that Experiment 3 (silent gap) produced significantly more correct identifications ($p < .05$) than did Experiment 1 (noise). A graphical representation of the data used in this analysis is shown in Fig. 3.

DISCUSSION

Since our last publication on phonemic restorations (Warren & Obusek, 1971), evidence has been reported from this laboratory indicating that perceptual synthesis leading to clear perception of missing sounds is not limited to speech. If certain rather rigid criteria are met, absent tones and noise bands may be heard clearly and may seem to continue for durations up to a few seconds. This "auditory induction" requires both acoustic context leading to the expectation of the missing sound and a louder sound of appropriate intensity and spectral composition (Warren, Obusek, & Ackroff, 1972). Three types of auditory induction have been described involving: (1) inducing and induced sounds of identical spectral composition (differing only in intensity); (2) sounds of differing spectral composition (e.g., a tone and a louder noise band); (3) phonemic restorations. This last type seems to correspond to a highly specialized form of auditory induction, in which linguistic rules enter into the synthesis of the restored sound.

Mislocalization of the extraneous sound inducing phonemic restorations occurs despite the direct involvement of this sound in the precisely localized processes leading to perceptual synthesis, indicating that

these processes are not directly accessible to the listener's attentional scrutiny or control. The mislocalization appears to be related to the general inability to locate the relative position of an item in an auditory sequence unless the item is incorporated directly within a verbal or melodic grouping (Warren, 1968; Warren, Obusek, Farmer, & Warren, 1969). The directional bias observed in the present study (that is, the preposing of the noise relative to its actual position) may reflect in part the additional processing time required for phonetic units near a phonemic restoration. In an earlier study, it was shown that, while Ss could not localize a click embedded within an otherwise intact phoneme, they did not prepose this extraneous sound as did a separate group hearing a longer extraneous sound replace the same phoneme in the otherwise identical sentence (Warren & Obusek, 1971). It seems that Ss listening to a sentence may store the auditory input in some partially processed form until confirmation of a particular verbal organization is achieved. Long storage may be required if a portion of a sentence is missing—especially if the prior context does not identify unambiguously the absent portion, requiring information occurring after the missing speech sound to complete the identification (as with the seven sentences used in the present study). The delay in completion of perceptual processing of the verbal signal near the noise burst would cause the noise to appear to occur with an earlier portion of the sentence, as was observed.

It is important to note that phonemic restoration, once achieved, produces phonemes perceptually indistinguishable from those actually present. Even being informed that a portion of the sentence was replaced by noise did not help our Ss in Experiment 2 to identify

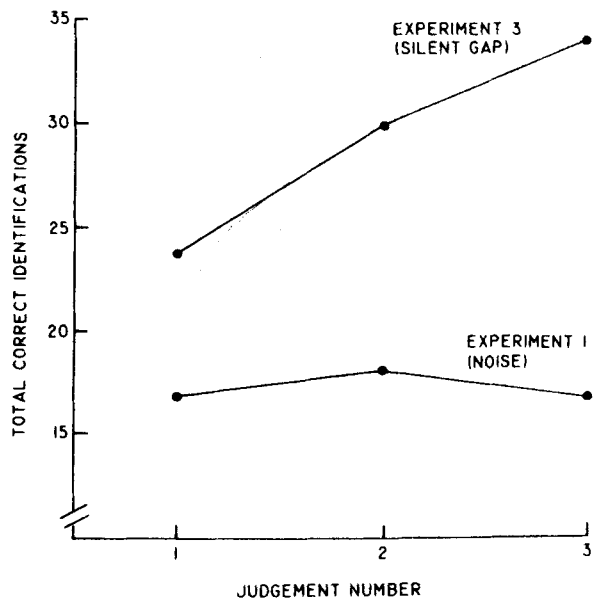


Fig. 3. Total number of correct identifications of missing phonemes in Experiments 1 and 3.

which speech sound they synthesized. Figure 2 shows that these listeners could not identify accurately the location of the noise (and the missing phoneme) even after two repetitions of the sentences. The Es and other staff and graduate students familiar with the experimental design still "heard" the sound consistent with context. The sound actually removed (for example, the /s/ of "sandwagon" in Sentence 1) was never reported (but only /b/ in "bandwagon" which was consistent with context).

Phonemic restorations are more than Gestalt-type closures or completion of gaps. If this filling-in were all, it would be anticipated that the silent gap would elicit responses not appreciably different from those with a noise-filled gap, but it was found that Ss could localize the silence (and identify the missing phoneme) more accurately. The relative ease of identifying the silent gaps seemed even more evident with informal testing of individuals who had training in listening to sentences analytically.

In the present study, context for phonemic restoration followed the deleted segment within the same word. However, some unpublished observations in this laboratory have indicated that the resolving context may be delayed for two or three, or even more, words following the ambiguous word fragment (Sherman, 1971). There is also evidence that the duration of the noise-filled gap must approximate the normal duration of the restored phoneme for successful phonemic restoration. Indeed, the phoneme restored in an ambiguous sentence can be changed by changing the length of the spliced-in noise burst. In the current experiments, the mispronounced deleted phonemes were carefully chosen to match the durations of the contextually correct phonemes.

Outside the laboratory, when employing phonemic restorations for comprehension under noisy conditions, several types of cues to the identity of the masked sound may be used on occasion: not only context and temporal matching, but transitional probabilities, coarticulation, spectral and intensive comparison of inducing and induced sounds, and perhaps even visual observation of the articulatory positions of the speaker's lips and jaws. Also, rather than a single extraneous sound in an otherwise clear sentence, as in the present study, phonemic restorations may occur often with multiple interruptions. It is of interest that Miller and Licklider (1950) interrupted a reading of PB ("phonetically balanced") word lists 10 to 15 times/sec with a louder noise and reported that "the listener feels that the speech is certainly more natural and probably more intelligible [p. 173]" than when the interruptions were left silent. However, they found that no actual improvement in intelligibility occurred. If we consider the noisy conditions as tending to elicit phonemic restorations, their lack of supplementary cues would halt the process short of perceptual synthesis. But if

interruptions by noise occurred along with the context provided by normal discourse, then intelligibility should be greater than when the interruptions consisted of unfilled silent gaps. Cherry and Wiley (1967) and Holloway (1970) described experiments relevant to this expectation in which the strongly voiced components in discourse were removed and either filled with noise or left as silent gaps separating the other components in the sentences. Intelligibility was found to be considerably greater when the noise was present. In an interesting unpublished study, Wiley (1968) reported an experiment resembling that of Miller and Licklider, except that Wiley used regularly interrupted discourse rather than regularly interrupted PB words. Wiley found that addition of noise to the regularly spaced gaps improved intelligibility, indicating that the published effect reported by Wiley and Cherry and by Holloway was not limited to the strongly voiced components of speech. In a paper delivered recently, Powers (1973) also reported that addition of noise to the silent intervals of regularly interrupted speech increases intelligibility. This curious increase in clarity of speech through the addition of noise becomes clarified conceptually if we consider that multiple phonemic restorations are induced in these sentences by the noise bursts.

The evidence available at this time suggests that a highly specialized form of auditory induction called phonemic restoration normally facilitates the comprehension of speech which is partially masked by intermittent extraneous sounds or fluctuating noise levels in the environment.

REFERENCES

- Cherry, C., & Wiley, R. L. Speech communication in very noisy environments. *Nature*, 1967, 214, 1164.
- Holloway, C. M. Passing the strongly voiced components of noisy speech. *Nature*, 1970, 226, 178-179.
- Miller, G. A., & Licklider, J. C. R. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 1950, 22, 167-173.
- Powers, G. L. Intelligibility of temporally interrupted speech with and without intervening noise. *Journal of the Acoustical Society of America*, 1973, 54, 300. (Abstract)
- Sherman, G. L. Phonemic restoration: An insight into the mechanisms of speech perception. Unpublished master's thesis, University of Wisconsin-Milwaukee, 1971.
- Warren, R. M. Relation of verbal transformations to other perceptual phenomena. Conference Publication No. 42, Institution of Electrical Engineers, London, 1968, Suppl. No. 1, 1-8.
- Warren, R. M. Perceptual restoration of missing speech sounds. *Science*, 1970, 167, 392-393.
- Warren, R. M. Identification time for phonemic components of graded complexity and for spelling of speech. *Perception & Psychophysics*, 1971, 9, 345-349.
- Warren, R. M., & Obusek, C. J. Speech perception and phonemic restorations. *Perception & Psychophysics*, 1971, 9, 358-363.
- Warren, R. M., Obusek, C. J., & Ackroff, J. M. Auditory induction: Perceptual synthesis of absent sounds. *Science*, 1972, 176, 1149-1151.
- Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusion of patterns other than speech or music. *Science*, 1969, 164, 586-587.
- Wiley, R. L. Speech communication using the strongly voiced components only. Unpublished PhD thesis, Faculty of Engineering, University of London, 1968.

(Received for publication August 4, 1973;
accepted March 4, 1974.)