Phonemic restoration of a stop consonant in a vowel-consonant-vowel syllable in speeded and nonspeeded tasks

Shigeaki Amano

*NTT Basic Research Laboratories,
3-9-11 Midori-cho, Musashino, 180 Japan*

(Received 13 August 1993)

Perceptual processing to integrate distributed consonant information was investigated by measuring phonemic restoration rates of a noise-replaced stop consonant within a vowel-consonant-vowel (VCV) syllable in speeded and nonspeeded tasks. Reaction times to phonemic restoration were also measured. Both speeded and nonspeeded perceptual processing were found to use the distributed information in almost the same way. The processing time was short when the restoration rate was high, but it was long when the restoration rate was low.

Keywords: Phonemic restoration, Phoneme detection, Reaction times, Speeded and nonspeeded tasks, Distributed consonant information

PACS number: 43. 71. An, 43. 71. Es

1. INTRODUCTION

Even if a short part of speech is replaced by noise, subjects usually hear the speech as if it were intact and had no missing part. This is known as phonemic restoration (Warren, 1970; Warren and Obusek, 1971). Phonemic restoration is affected by top-down information, such as sentential information (Warren and Sherman, 1974), lexical information (Samuel, 1987), attention (Samuel, 1991; Samuel and Ressler, 1986), and the native language of the subject (Kashino, van Wieringen and Pols, 1992). It is also affected by bottom-up information, such as the acoustic similarity between the phoneme and the noise (Samuel, 1981a, 1981b). But the most important factor contributing to phonemic restoration is that a number of cues for each phoneme are distributed in the speech waveforms (*e.g.*, Kashino, 1992; Liberman, 1970). Many studies have shown that distributed cues, such as coarticulatory information, contribute to phoneme perception (Amano, 1993; Fowler, 1984;

Fowler and Smith, 1986; Martin and Bunnell, 1981; Martin and Bunnell, 1982; Mills, 1980). Phonemic restoration greatly depends on the time-distributed information preceding or following the noise-replaced part. For example, Kashino (1992) showed that a noise-replaced stop consonant in a Japanese vowel-consonant-vowel syllable was restored from the phonemic information before and after its closure part.

Almost all experiments on phonemic restoration have used a nonspeeded task in which subjects had some time before giving their response. A typical example of a nonspeeded task was for subjects to listen to a noise-replaced stimulus and then write down what they perceived, as Kashino (1992) did. There was no time constraint on the task. So, it is possible that the subjects could restore a noise-replaced phoneme because they could precisely and carefully analyze the stimulus before they responded. This raises the question of whether subjects can restore a noise-replaced phoneme in a speeded task, in which they must respond im-

mediately.

However, there are no published studies on how a time constraint affects to phonemic restoration; a very few studies on speech perception have raised the question of the difference between speeded and nonspeeded perceptual processing. One is the study by Tyler and Wessels (1985) of gating task in which the perception of words spoken in a sentence was explored. They compared the results of a speeded gating task with those of a nonspeeded gating task and found that the results were very similar. This indicates that a time constraint has only a small effect on spoken word perception.

Is their finding on the time constraint also applicable to phoneme perception? More specifically, does the same phonemic restoration occur in speeded and nonspeeded tasks and is the same distributed information used in both speeded and nonspeeded phonemic restoration?

To answer these questions, restoration rates for a noise-replaced phoneme were measured when the subjects responded quickly and when they had time to consider their response. Reaction times for phonemic restoration were also measured when the subjects responded quickly, in order to obtain further understanding of speeded phonemic restoration.

2. EXPERIMENT 1

The aim of Experiment 1 was to replicate the experiment of Kashino (1992) and to obtain nonspeeded restoration data for comparison with speeded data.

2.1 Method

2.1.1 Subjects

The subjects were four female native Japanese speakers (23–34 years old) with normal hearing ability. They are employees at NTT Laboratories and are highly trained in evaluating the quality and intelligibility of speech.

2.1.2 Materials

The original stimuli were vowel-consonant-vowel (VCV) Japanese syllables. The first vowel was /a/, /e/, or /o/. The consonant was of either /p/ or /k/. The second vowel was /a/, /e/, or /o/. Thus, there were 18 VCV syllables. A native Japanese female speaker pronounced these original stimuli at an approximately constant rate. The accent patterns of the stimuli were “flat type,” which means that the pitch of the first vowel was

Table 1 Mean duration of parts of original stimuli (ms). Standard deviation is shown in parenthesis.

Consonant	First vowel	Closure part	Burst part	Second vowel
/k/	123 (8)	138 (11)	27 (5)	197 (13)
/p/	129 (12)	139 (10)	10 (3)	214 (23)

low and that of the second vowel was high. The stimuli were recorded at a 16-kHz sampling rate with 16-bit A/D conversion in a soundproof booth. The signal-to-noise ratio was 56.5 dB. The power of each stimulus was digitally normalized on a computer after recording. The mean duration of the parts of the original stimuli are shown in Table 1.

In this paper, C1 and C2 represent the consonants indicated by the pre- and post-closure parts, respectively. The transition at the end of the first vowel contributes to C1, while the burst of the stop consonant and the transition from the burst to the second vowel contributes to C2. C1 and C2 were arranged in the stimulus conditions:

1. Under the Normal condition, the original stimuli are used as stimuli. C1 and C2 would be the same phoneme if there is no noise replacement which is described later.

2. Under the V-C2V condition, stimuli are created by digitally exchanging the first vowel in the original stimuli with a different vowel that has the same pronunciation. This new vowel is pronounced independently by the same speaker. The exchange point is the end of the first vowel. The duration of the independently pronounced vowel is matched to that of the original first vowel by shaping the end of its waveform with a half-cycle cosine window (40 ms). The power of the independently pronounced vowel is digitally normalized on a computer before this operation. Under this condition, information for C2 exists but information for C1 does not, even though information for the first vowel itself exists.

3. Under the VC1-C2V condition, stimuli are created by digitally exchanging the first vowels of two stimuli that share the same first and second vowels. That is, the first vowel followed by /k/ and the other first vowel followed by /p/ are exchanged to construct two new stimuli. The exchange point is the end of the first vowel. Under

this condition, C1 and C2 would be different phonemes.

4. Under the C2V condition, stimuli are created by digitally deleting the first vowel and closure part of the original stimuli. Under this condition, information for C1 and information for the first vowel does not exist; only information for C2 exists.

Under these four conditions, every stop consonant in the stimuli was replaced by white noise for durations of 0, 10, 30, 50, or 70 ms. The onset of the noise was synchronized with the onset of the burst of the stop consonant. The power of the noise was 5 dB greater than the average power of the stimulus. As shown in Table 1, 70-ms noise was long enough to replace the whole burst part and the beginning part of the second vowel, which contributes to C2. Therefore, C2 with 0-ms noise replacement (*i.e.*, C2 without noise replacement) and C2 with 70-ms noise replacement would be different phonemes. In other words, the original C2 at 0-ms noise replacement would be changed to a non-original C2 because of the noise replacement. There were 360 stimuli in total.

2.1.3 Procedure

The stimuli were presented by a personal computer with a DMA board controlling a 16-bit D/A converter. The D/A output was filtered by an 8-kHz, 135-dB/octave low-pass filter and then amplified. The stimuli were diotically presented to the subject through headphones (SR-APRO, STAX) at 71 dB SPL in a listening booth with 24-dB (A) background noise. The stimulus order was randomized for each subject by the personal computer. For each trial, a pure tone (1,000 Hz, 150 ms) was presented first, followed by a 1,000-ms foreperiod, and then the stimulus. The interval between successive trials was about 3 s. The subjects were instructed to write down what they heard in Japanese characters (hiragana). The subjects did four blocks of experiments. One block consisted of 360 trials after 36 training trials. There was a 3-min break after every 100 trials.

2.2 Results and Discussion

Figure 1a shows the rates of restoration of the original C2; that is, the subject reported /k/ when the burst and the following transition originally indicated /k/, or the subject reported /p/ when the burst and the following transition originally indi-

Fig. 1 Nonspeeeded restoration rate of a noise-replaced consonant as a function of noise duration when (a) the original C2 was restored, and (b) the non-original C2 was restored.

cated /p/. The restoration rates of the original C2 decreased as the noise duration increased, indicating that information for the original C2 was lost due to the noise replacement.

Figure 1b shows the rates of restoration of the non-original C2; that is, the subject reported /k/ when the burst and the following transition originally indicated /p/, or the subject reported /p/ when the burst and the following transition originally indicated /k/. The restoration rates of the non-original C2 increased as the noise duration increased, indicating that information for the non-original C2 increased due to the noise replacement. These two cases show that information at the burst and the following transition, whether it contributes to the original or non-original C2, contributes to phonemic restoration.

In addition to this contribution, information for C1 may also contribute to phonemic restoration because restoration rates are different under Normal, V-C2V, VC1-C2V, and C2V conditions (Figs. 1a and 1b). To see how information for C1 contributes to phonemic restoration, statistical tests of percentage difference were conducted between every pair of stimulus conditions (*i.e.*, Normal, V-C2V, VC1-C2V, C2V) for both original and non-original C2 restoration. When the original C2 was restored, there were no significant differences between any pair of stimulus conditions at 0- or 10-ms noise duration. But there were 1% significant differences between all pairs of stimulus conditions at 30-, 50, and 70-ms noise duration, except between V-C2V and C2V conditions at 30- and 50-ms

noise duration. The percentage difference test was not conducted at 0- or 10-ms noise durations when the non-original C2 was restored, because there was very little data for these noise durations. There were 1% significant differences between all pairs of stimulus conditions at 30-, 50-, and 70-ms noise duration, except between V-C2V and C2V conditions at 30-ms noise duration.

These results confirm that information for C1 contributes to phonemic restoration. And when C1 matched the original C2 under the Normal condition or when it matched the non-original C2 under the VC1-C2V condition, C1 supplemented C2 and the restoration rate was high. But when C1 did not match the original C2 under the VC1-C2V condition or when it did not match the non-original C2 under the Normal condition, C1 competed with C2 and the restoration rate was low.

These results are almost the same as Kashino's (1992) results, but with one difference. In Kashino's experiment, the restoration rate of the original C2 under the C2V condition was almost the same as that under the VC1-C2V condition. In my Experiment 1, however, the restoration rate of the original C2 was greater under the C2V condition than that under the VC1-C2V condition. The results of Experiment 1 are more reasonable, however, because C1 competes with C2 under the VC1-C2V condition, which must make the restoration rate lower; there is no competition, however, under the C2V condition.

Because information for C1 did not exist under either C2V or V-C2V conditions, the restoration rates could be the same. However the restoration rate under the V-C2V condition was higher than that under the C2V condition when the original C2 was restored. One reason for this is that information about the closure part was available for phonemic restoration under the V-C2V condition, but not under the C2V condition. Another possibility is that information very similar to C2 was created by the digital cutting operation at the end of the first vowel.

Although there are a few differences from Kashino's results, the results of Experiment 1 basically support his findings.

3. EXPERIMENT 2

The aim of Experiment 2 was to obtain speeded restoration rates and restoration reaction times.

3.1 Method

3.1.1 Subjects

The subjects were the same as in Experiment 1. All subjects were well trained for a reaction-time experiment.

3.1.2 Materials and procedure

The same stimuli and apparatus as in Experiment 1 were used. The subjects performed phoneme detection individually. The target phoneme was the stop consonant in the VCV stimulus. The target was /k/ in half the experiments and /p/ in the other half, using the same stimulus set.

The subjects were instructed to respond as quickly and accurately as possible by pressing one of two keys. They pressed one key with their first finger when they heard the target phoneme and the other key with their second finger when they perceived that the stimulus did not contain the target phoneme. The keys were connected to the personal computer by a parallel interface board, and responses were recorded automatically. The restoration rate was obtained from the number of key responses to the target phoneme. The reaction times for the target phoneme were measured from the burst onset of the stop consonant. One block consisted of 360 trials after 70 training trials. There was a 3-ms break after every 70 trials. The subject did four blocks for both targets.

3.2 Results and Discussion

Responses with a reaction time of more than 1,000 ms or less than 0 ms were excluded as errors. The error rate was 0.12%.

Figure 2a shows the restoration rates of the

Fig. 2 Speeded restoration rate of a noise-replaced consonant as a function of noise duration when (a) the original C2 was restored, and (b) the non-original C2 was restored.

original C2; that is, the subject detected the target phoneme /k/ when the burst and the following transition originally indicated /k/, or the subject detected the target phoneme /p/ when the burst and the following transition originally indicated /p/. The restoration rates of the original C2 decreased as the noise duration increased. Figure 2b shows the restoration rates of the non-original C2; that is, the subject detected the target phoneme /k/ when the burst and the following transition originally indicated /p/, or the subject detected the target phoneme /p/ when the burst and the following transition originally indicated /k/. The restoration rates of the non-original C2 increased as the noise duration increased. These tendencies are the same as in Experiment 1, confirming that information for C2 also contributes to speeded phoneme restoration.

To see how information for C1 contributes to phonemic restoration, statistical tests of the percentage difference were conducted between every pair of stimulus conditions for both the original and the non-original C2 restoration. When the original C2 was restored, there were significant differences between V-C2V and C2V conditions at 0-ms noise duration ($z=3.19$, $p<0.01$), between Normal and VC1-C2V conditions at 10-ms noise duration ($z=4.03$, $p<0.01$), between Normal and C2V conditions at 10-ms noise duration ($z=12.5$, $p<0.01$), between V-C2V and C2V conditions at 10-ms noise duration ($z=4.23$, $p<0.01$), and between VC1-C2V and C2V conditions at 10-ms noise duration ($z=2.38$, $p<0.05$). There were 1% significant differences between all pairs of stimulus conditions at 30-, 50-, and 70-ms noise durations, except between V-C2V and C2V conditions at 30-ms noise duration. The percentage difference test was not conducted at 0- and 10-ms noise durations when the non-original C2 was restored, because there was very little data for these noise durations. There were 1% significant differences between all pairs of stimulus conditions at 30-, 50-, and 70-ms noise durations, except between V-C2V and C2V conditions at 30-ms noise duration. These results showed that information for C1, as well as C2, contribute to speeded phonemic restoration.

Statistical tests of percentage differences between speeded and nonspeeded restoration rates were conducted for each stimulus condition and noise duration for both the original and non-original C2

restoration. When the original C2 was restored, there were significant differences at 0-ms noise duration under the Normal condition ($z=2.13$, $p<0.05$), VC1-C2V condition ($z=3.04$, $p<0.01$), and C2V condition ($z=2.13$, $p<0.01$). There were also significant differences at 10-ms noise duration under the VC1-C2V condition ($z=2.11$, $p<0.05$) and C2V condition ($z=3.93$, $p<0.01$). When the non-original C2 was restored, there were significant differences at 0-ms noise duration under the Normal condition ($z=2.66$, $p<0.01$), VC1-C2V condition ($z=2.46$, $p<0.05$), C2V condition ($z=3.50$, $p<0.01$), and V-C2V condition ($z=2.46$, $p<0.05$); at 10-ms noise duration under the VC1-C2V condition ($z=4.22$, $p<0.01$), C2V condition ($z=3.02$, $p<0.01$), and V-C2V condition ($z=2.07$, $p<0.05$); and at 30-ms noise duration under the C2V condition ($z=2.27$, $p<0.05$).

In spite of these point-to-point differences, no significant difference between speeded and nonspeeded restoration rates was observed ($t=0.20$, $p>0.05$) when a paired comparison was conducted using all data points in Figs. 1 and 2. The speeded phonemic restoration rate is almost the same as the nonspeeded rate, indicating that distributed consonant information, such as for C1 and C2, contributes to both speeded phonemic restoration and nonspeeded phonemic restoration.

Figure 3 gives the mean reaction times for restored consonants as a function of noise duration. The analysis of variance of reaction times was conducted regarding subjects as a random factor. There were significant interactions between the noise duration and the stimulus condition both when the

Fig. 3 Mean reaction times (RT) to a noise-replaced consonant as a function of noise duration when (a) the original C2 was restored, and (b) the non-original C2 was restored.

original C2 was restored [$F(12, 3585)=1.83$, $p < 0.05$] and when the non-original C2 was restored [$F(12, 900)=1.97$, $p < 0.05$]. Therefore, the analysis of variance for the simple main factor of the stimulus condition was conducted at every noise duration. When the original C2 was restored, the factor of the stimulus condition was significant for all noise durations [at 0 ms, $F(3, 1102)=32.77$, $p < 0.01$; at 10 ms, $F(3, 1038)=40.08$, $p < 0.01$; at 30 ms, $F(3, 561)=13.56$, $p < 0.01$; at 50 ms, $F(3, 427)=11.81$, $p < 0.01$; and at 70 ms, $F(3, 395)=15.71$, $p < 0.01$]. Multiple comparisons with Tukey's HSD method revealed 5% significant differences between all pairs of stimulus conditions, except for the V-C2V and C2V conditions at every noise duration, the Normal and C2V conditions at 50-ms noise duration, and the VC1-C2V and C2V conditions at 70-ms noise duration.

These results show that reaction times were significantly faster under the Normal condition than under other conditions, significantly slower under the VC1-C2V conditions than under other conditions, and almost the same under the V-C2V and C2V conditions. C1 under the Normal condition facilitates restoration, but C1 under the VC1-C2V condition inhibits restoration when the original C2 is restored.

The opposite occurred when the non-original C2 was restored. The analysis of variance for the simple main factor of the stimulus condition was conducted at 30-, 50, and 70-ms noise durations, but not at 0- or 10-ms noise durations, because there was very little data at these noise durations. The factor of the stimulus conditions was significant at 50-ms noise duration [$F(3, 303)=6.93$, $p < 0.01$] and at 70-ms noise duration [$F(3, 316)=17.2$, $p < 0.01$]. Multiple comparisons with Tukey's HSD method revealed 5% significant differences between all pairs of stimulus conditions at 50-ms noise duration, except for the V-C2V and C2V pair conditions and the VC1-C2V and C2V pair conditions. Multiple comparisons also revealed 5% significant differences between all pairs of stimulus conditions at 70-ms noise duration, except for the V-C2V and C2V pair conditions.

These results show that reaction times were significantly slower under the Normal condition than under other conditions, significantly faster under VC1-C2V conditions than under other conditions, and almost the same under V-C2V and C2V condi-

tions. C1 under the VC1-C2V condition facilitates restoration, but C1 under the Normal condition inhibits restoration when the non-original C2 is restored.

The contribution of C1 revealed by the reaction times is much the same as in the restoration rates. That is, when C1 matched the original C2 under the Normal condition or when it matched the non-original C2 under the VC1-C2V condition, C1 supplemented C2 and the reaction time was faster. However, when C1 did not match the original C2 under the VC1-C2V condition or when it did not match the non-original C2 under the Normal condition, C1 competed with C2 and the reaction time was slower. The results for reaction times, as well as restoration rates, imply that both speeded and nonspeeded phonemic restoration are based on the same perceptual processing.

4. GENERAL DISCUSSION

The restoration rates of speeded and nonspeeded phonemic restoration were almost the same, and differences in reaction times to speeded phonemic restoration under various stimulus conditions were consistent with differences in restoration rates. These results show that speeded and nonspeeded phonemic restoration are the same: they both use distributed consonant information. In other words, even if subjects are given more time to analyze the acoustic information, they do not extract any more information from the speech waveform than that they have already obtained with speeded analysis.

This indicates that fast phonemic restoration is an unconscious automatic process for determining which phoneme exist at noise-replaced part, and that subjects do not consciously control the process. However, the present study only investigated the case in which there was only VCV information. Previous studies have shown that phonemic restoration is also affected by sentence information (Warren and Sherman, 1974) or visual cues that indicate to which phoneme attention should be paid (Samuel, 1991; Samuel and Ressler, 1986). Although those studies used a nonspeeded task, top-down information was given before stimulus presentation. Therefore, it can be inferred that if such information were also available in speeded phonemic restoration before stimulus presentation, phonemic restoration would be affected.

Even after stimulus onset, if top-down informa-

tion is available, at least before the noise-replaced part, it can affect phonemic restoration. Lexical information is one such case. Samuel (1987) found that lexical information improves the phonemic restoration rate the nearer the noise-replaced phoneme is to the end of the word. While he used a nonspeeded task, lexical facilitation has also been obtained by some studies for a speeded task, *i.e.*, phoneme detection (*e.g.*, Amano, 1993; Marslen-Wilson, 1984). Although no noise-replacement was conducted in those studies, they did indicate that lexical information is effective in speeded phonemic restoration.

The reaction times for lexical and nonlexical items in Experiment 2 were analyzed, but no significant lexical effect was found. This was probably because the stimuli were only two mora long. In Samuel's (1987) study and in phoneme detection studies (Amano, 1993; Marslen-Wilson, 1984), the lexical effect was smaller when the phoneme was near the beginning of the word. Thus, if longer stimuli are used and the noise-replaced part is near the end of the word, a lexical effect should be observed even in speeded phonemic restoration.

In summary, the present study systematically investigated speeded and nonspeeded perception of phonemic restoration with no top-down information. Perception was almost the same, indicating that both use the same distributed phoneme information. The restoration rate was high and the reaction time was fast when the distributed phoneme information was consistent, but the restoration rate was low and the reaction time was slow when the distributed phoneme information was inconsistent.

ACKNOWLEDGEMENTS

I gratefully thank Kazuhiko Kakehi for his advice on the experiments and his useful comments on an earlier version of this manuscript.

REFERENCES

- Amano, S. (1993). "Effects of lexicon and coarticulation on phoneme perception," *J. Acoust. Soc. Jpn. (E)* **14**, 91-97.
- Fowler, C. A. (1984). "Segmentation of coarticulated speech in perception," *Percept. Psychophys.* **36**, 359-368.
- Fowler, C. A. and Smith, M. R. (1986). "Speech perception as 'vector analysis': An approach to the problems of invariance and segmentation," in *Invariance and Variability in Speech Processes*, J. S. Perkell and D. H. Klatt, Eds. (LEA, Hillsdale), Chap. 6, p. 123.
- Kashino, M. (1992). "Perception of Japanese intervocalic stop consonants based on the distributed cues in pre- and post-closure portions," *J. Acoust. Soc. Jpn. (J)* **48**, 76-86 (in Japanese).
- Kashino, M., van Wieringen, A., and Pols, L. C. W. (1992). "Cross-languages differences in the identification of intervocalic stop consonants by Japanese and Dutch listeners," *Proc. Int. Conf. Spoken Language Processing*, 1079-1082.
- Lieberman, A. M. (1970). "The grammars of speech and language," *Cognit. Psychol.* **1**, 301-323.
- Marslen-Wilson, W. D. (1984). "Function and process in spoken word recognition—A tutorial review," in *Attention and Performance*, Vol. 10, H. Bouma and D. G. Bouwhuis, Eds. (LEA, London), Chap. 8, p. 125.
- Martin, J. G. and Bunnell, H. T. (1981). "Perception of anticipatory coarticulation effects," *J. Acoust. Soc. Am.* **69**, 559-567.
- Martin, J. G. and Bunnell, H. T. (1982). "Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences," *J. Exp. Psychol. Hum. Percept. Perform.* **8**, 473-488.
- Mills, C. B. (1980). "Effects of the match between listener expectancies and coarticulatory cues on the perception of speech," *J. Exp. Psychol. Hum. Percept. Perform.* **6**, 528-535.
- Samuel, A. G. (1981a). "Phonemic restoration: Insights from a new methodology," *J. Exp. Psychol. Gen.* **110**, 474-494.
- Samuel, A. G. (1981b). "The role of bottom-up confirmation in the phonemic restoration illusion," *J. Exp. Psychol. Hum. Percept. Perform.* **7**, 1124-1131.
- Samuel, A. G. (1987). "Lexical uniqueness effects on phonemic restoration," *J. Mem. Lang.* **26**, 36-56.
- Samuel, A. G. (1991). "A further examination of attentional effects in the phonemic restoration illusion," *Q. J. Exp. Psychol.* **43A**, 679-699.
- Samuel, A. G. and Ressler, W. H. (1986). "Attention within auditory word perception: Insights from the phonemic restoration illusion," *J. Exp. Psychol. Hum. Percept. Perform.* **12**, 70-79.
- Tyler, L. K. and Wessels, J. (1985). "Is gating an on-line task? Evidence from naming latency data," *Percept. Psychophys.* **38**, 217-222.
- Warren, R. M. (1970). "Perceptual restoration of missing speech sounds," *Science* **167**, 392-393.
- Warren, R. M. and Obusek, C. J. (1971). "Speech perception and phonemic restorations," *Percept. Psychophys.* **9**, 358-362.
- Warren, R. M. and Sherman, G. L. (1974). "Phonemic restorations based on subsequent context," *Percept. Psychophys.* **16**, 150-156.