Sentential Context Effect on Phoneme Restoration in Fluent Speech
—Is it necessarily a post-perceptual decision bias?

Eiichi YUBUNE
Toyo University

It has been suggested that lexical context seems to play a robust and direct role in phoneme restoration (more generally, perception) and the effect is considered to be a true feedback mechanism, whereas sentential context does not directly influence phoneme level processing and the effect is merely a post-perceptual decision bias. The experiment reported here examines how word-final ambiguous phonemes as artifacts of phonological processes such as consonant deletion and devoicing are restored by utilization of sentential context. The results show the possibility that sentential context can directly influence phoneme level processing. The data also provide evidence which supports the view that there is a level of phoneme representation in spoken word recognition as has been claimed in TRACE model, a connectionist model of speech perception and word recognition. I suggest that the nature of the role of context is not necessarily dichotomous between the purely true feedback and the post-perceptual decision bias, and is dependent on the degree of the constraint intrinsic to context.

1. INTRODUCTION: A Critical Review
1.1 Sentential Context Effect on Phoneme Processing

Phoneme restoration phenomenon is a powerful auditory illusion by which listeners hear parts of a word that are not really there. In earlier studies of phoneme restoration (Warren, 1970; Warren and Obusek, 1971), segments of words (phonemes) were replaced by an extraneous sound such as a cough. Most subjects tended to report hearing an intact word and mislocalize the noise, suggesting that they had restored the missing phoneme.

Samuel (1981) has reported a series of studies on phoneme restoration using signal detection methodology to separate actual perceptual effects from decision bias. In these experiments, he used a second type of replacement noise which was merely added to the appropriate phoneme in addition to the usual replaced items, that is, a sound replaced by white noise. The utility of this type of phoneme restoration task is that the added or replaced technique allows subjects’ biases ($\beta$) to be separated from sensitivities ($d'$). If sentential context (semantic or syntactic) is directly influencing the phoneme level processing, discriminability of replaced and added sounds in target words should decrease when the target words are predictable. Samuel found that the discriminability actually increased rather than decreased in a predictable sentential context. In contrast, subjects tended to report hearing target words as intact rather than replaced when they occurred in a contextually appropriate sentence. These results suggest that sentential context may affect the post-perceptual decision stage, biasing listeners to report utterances as intact, but may not directly contribute to
phonetic processing.

A different task used to investigate sentential context effects on speech perception is the identification function shift paradigm. Connine (1987) investigated semantic context effects on identification of speech sound, using tokens from voicing series (e.g., dent—tent) which were embedded in sentence contexts semantically biased toward the voiced or voiceless endpoints. Subjects labeled ambiguous sounds (stimuli in the category boundary region) such that a semantically consistent word was formed. For example, ambiguous stimuli between the dent—tent continuum tended to be identified as /d/ in the context She drives a car with the and as /t/ in the context She saw the show in the. However, identification of the endpoint stimuli was not influenced by the context (also reported in Miller et al. (1984)). She also found that reaction times for contextually appropriate responses to ambiguous stimuli were not faster than contextually inappropriate responses. In fact, responses for ambiguous stimuli were generally slower than those for the endpoint stimuli. Only at the endpoints of the continuum, a reaction time advantage for context consistency was gained.

These reaction time data are consistent with the TRACE model of speech perception (McClelland and Elman, 1986), in which the feedback from sentence level context to perceptual processing is assumed to be a time-constrained process. Therefore, since sentence context needs time to activate its effect, the influence of higher semantic/syntactic context on perceptual level processing will be greatest when stimuli are ambiguous. When stimuli are unambiguous, the bottom-up data will be sufficient and fast enough so that no time is left for sentential context effect to start activating.

Therefore it appears so far that sentential context effect does not pass down to the phoneme level processing.

1.2 Effect of Lexical Context

Lexical effect on phoneme perception is also worth discussing here since if higher level contexts were to influence sublexical processing they would inevitably pass through the lexical level processor, because in normal speech sounds are realized to form words. And it is important to know what is the underlying nature of the difference between these two different loci of context effects by comparing them with each other. In contrast to sentence context effects, it seems that the effect of lexical status in phoneme perception is rather straightforward. A number of studies have argued that lexical context directly influences perceptual processing.

First, in Samuel’s (1981) study discussed earlier, he also conducted an experiment in which lexical effect was investigated with phoneme restoration using the same signal
detection methodology. The results supported a true feedback evidence between lexical and phoneme level processing. Discriminability of replaced or added noise with intact sound was decreased more when these manipulated sounds occurred in a word than when in a non-word, indicating that lexical knowledge was perceptually restoring the manipulated sound.

Ganong (1980) reported that the point of listener’s identification shift in a speech continuum between two sounds, say /d/ to /t/, differs depending on which end of the continuum forms a word. Listeners tend to hear /d/ if the sound is followed by *ash* whereas /t/ before *ask*. This is called the Ganong effect. Connine and Clifton (1987) used this identification paradigm to investigate the direct effect of lexical context on speech perception. They reported that subjects tended to label ambiguous stimuli so as to form a word. For example, stimuli in the midrange of the continuum between /t/ and /d/ tended to be heard as /t/ when they occurred in the context *TYPE-DYPE* while /d/ was heard in the context *DICE-TICE*. In contrast, at the end of the continuum, no systematic identification responses were found. In the reaction time analysis for the category boundary region and continua endpoints, a reaction time advantage was found only at the category boundary although no such advantage was found at the endpoints.

The pattern of the reaction times data presented here together with Samuel’s phoneme restoration study is regarded as evidence for direct contribution of lexical context to interpretation of acoustic input. This is contrary to the results from the studies on sentential context effect detailed in the previous section. In other words, reaction times were faster at the endpoints of the speech continua in a predictable word in a sentence and reaction times for ambiguous region were slower regardless of the predictability from the sentential context, suggesting that sentential context may support only post-perceptual decisions, that is, decision bias. On the other hand, lexical effect improves reaction times for ambiguous stimuli while no advantage is evident for canonical sound, suggesting that lexical information plays a direct and robust role in speech processing, that is, true perception.

So far, we have reached rather antagonistic conclusions; lexical context seems to play a robust and direct role in phoneme restoration (more generally, perception) and is considered to be a true feedback mechanism, whereas sentential context seems not to influence directly to phoneme level processing and is considered to be just a post-perceptual decision bias.

1.3 Phonological Variation and Abstract Phoneme Representation

Perception of speech sound is greatly complicated by various processes of phonological variation and the phonetic realization of words varies according to the
phonological context in which they occur. The topic has recently been extensively investigated by Marslen-Wilson and his colleagues (Gaskell and Marslen-Wilson, 1996; Marslen-Wilson et al, 1995; Warren and Marslen-Wilson, 1987, 1988). According to the underspecification theory in phonological study (Archangeli, 1988; Pulman and Hepple, 1993) and a series of studies by Marslen-Wilson et al., in normal speech potential ambiguities can be resolved by different levels of constraints—phonological, lexical and sentential. For instance, Marslen-Wilson et al (1995) provide the evidence that lexical representations are abstract and probably underspecified, and that the treatment of variation is highly context-sensitive. Let us take the ambiguous velar /k/, for example:

Phonological level: in [stelk] as in *steak dinner*, the ambiguity is resolved phonologically because the following segment /d/ does not have velar place.

Lexical level: in [swik] as in *sweet girl*, the ambiguity is resolved by the absence of a lexical item such as *sweek, sneak* and so on.

Sentential level: in [elk] as in *There were eight kids at the party*, the ambiguity is resolved through the syntactic and semantic knowledge that *ache kids* is not acceptable and *eight kids* is most appropriate.

Thus, sensitivity to co-articulation may well be discussed in the framework of abstract phoneme representation, and in some cases an ambiguous segment due to co-articulation will be disambiguated only through the phonological sensitivity to co-articulation, but in other cases it may only be resolved by way of higher levels of contexts. And, among these constraints, the role of sentential context is the most relevant issue to the present study.

In addition to co-articulation processes, there are many other types of phonetic realization in which a phoneme is altered in terms of its voicing feature. It is well described in the literature of phonetics that voiced consonants tend to be voiceless when they occur in word-initial or word-final position. Specifically for word-initial plosives, however, the voicing cue is provided as the VOT value so that it is relatively easy to perceive the voice opposition. For word-final plosives as well as fricatives, however, the possible cues for the voicing may be the quality and the quantity of the preceding vowel. In other words, the final voiceless consonants are distinguished by a comparatively short vowel with tight voicing and glottalization; and a relatively long vowel with full voicing preceding a voiced consonant (Brown, 1974). Although these
voicing cues may be reliable in isolated words, they may not be so evident in various phonological contexts in fluent speech. Thus, for example, *peace* and *peas* are less distinguishable in fast speech and other contextual information becomes a reliable cue; *peace* is more likely in the context *keep the peace* and *peas* in the context *eat the peas*. In this case, too, an abstract phoneme representation is being used to restore the altered phonemes by utilizing a higher-level context.

1.4 Restoration of Missing Phonemes

Samuel (1997) conducted an experiment in which he tried to investigate restoration effect by using an adaptation paradigm. Subjects were asked to make phonemic judgements about stimuli on speech continua such as /bi/ to /di/ after repeated presentation of adaptation words (e.g., *academic, armadillo*). When adaptation words with /d/ embedded were presented repeatedly in the adaptation phase, a significant shift in judgements of hearing /di/ was found in the subsequent phonemic judgement phase. He also compared this adaptation effect on intact adaptors with words containing a target phoneme replaced by noise. The results show a significant adaptation effect even to such manipulated phonemes. However, when the critical phoneme was simply deleted from words (e.g., *arma illo*), no such effect was found even though such a word was actually recognizable, suggesting that the restoration effect of this kind seems to be available only under the condition in which there is some replaced sound.

Then, what will be the case when it comes to the restoration of deleted phonemes acting purely on phonological rules? We can assume that such phonemes are more restorable when they occur in a word-medial position (e.g., /p/ as in *topmost* or /d/ as in *standstill*) through the underspecification rule or lexical knowledge. The problem then arises when such deletion occurs in a word which still is a lexical entry without that critical phoneme. For example, a past-tense or past-participle morpheme /d/ may often be deleted in a context such as *grabbed the doll*. Likewise, /t/ in *passed* will be deleted in a context such as *passed the exam*. Listeners would often garden-path for *grab the doll* or *pass the exam* without any appropriate sentential context such as *She was pleased with the present and grabbed the doll* or *He might have passed the exam*. In these circumstances, sentential context appears crucial for restoring the target phonemes. If this is the case, our interest should be whether this sentential context effect in restoring such particular phonemes is a true perceptual process or merely an artifact of post-perceptual decision.

So far, we can tentatively conclude that an ambiguous segment due to co-articulation or other phonological processes in fluent speech are restored as an appropriate abstract phoneme in listeners’ minds, guided by different levels of
constraints. If so, we can return to the contradictory view of the roles of lexical and sentential contexts discussed in earlier sections. In other words, will restoration of ambiguous or deleted phonemes due to real phonological processes in fluent speech retrace exactly the same processes we have specified, namely, a true feedback from the lexical level or a post-perceptual decision bias from the sentential level? In fact, studies on sentential context effect in phoneme restoration and perception so far have never incorporated the nature of the co-articulation systematically to the analyses of their target sounds. Tokens used in these experiments were just manipulated by a speech analyzer to produce speech continua or to add or replace the target segments with noise. Moreover, in Connine’s (1987) experiment all target phonemes were word-initial, and in Samuel’s (1981) series of experiments manipulated phonemes did not undergo normal co-articulation processes in fluent speech.

The experiment reported here examines how word-final ambiguous phonemes as artifacts of various phonological processes such as deletion and voice neutralization in word-final will be restored by utilization of sentential context. More importantly, the experiment helps to explore the way in which sentential context influences the perception of target phonemes used in the experiment—is it post-perceptual after all, as we have seen? Or is there any condition under which sentential context can directly influence the perception of these particular phonemes?

2. EXPERIMENT

The basic hypothesis of the present study is that there is a level of phoneme representation in language processing as can be seen in the TRACE model of speech perception, and listeners use this abstract representation to restore words (which, consequently, are also abstract) in a stretch of speech. According to the conditions for true perception discussed earlier, in predictable context if listeners actually hear (or restore) a certain phoneme immediately which they would not hear without any appropriate sentential contexts, then this can be called perception (a true feedback mechanism). On the other hand, if listeners actually hear an inappropriate phoneme in light of the context and they change the decision so that the phoneme may be consistent with the sentential context, then this is called a post-perceptual decision bias.

In the experiment, I used a phoneme judgement task in which I asked subjects to respond to the question of whether they have heard a target phoneme as quickly and intuitively as possible immediately after they heard each stimulus. I strictly required subjects to report whether they actually heard the target phoneme or not, even if they were able to recognize an appropriate word in a stimulus and consequently an appropriate phoneme.
There are two types of stimuli in the present experiment: one is for pairs of words which differ only in voice opposition of the word-final consonant (Experiment 1) and the other is for a series of past participles of regular verbs which end with /d/ or /t/ (Experiment 2). The word-final target phonemes in both types of stimuli are ambiguous in different ways; for all word pairs the target sounds tend to be heard as voiceless and for the verbs the target sounds tend to be heard as deleted. However, if correct phonemes are reported to be actually heard immediately in a predictable context, that is, if the carrier words are predictable, then we may expect listeners to perceive the critical phoneme, suggesting that sentential context is directly influencing the perceptual process.

2.1 Materials
I created stimuli for both Experiment Unit 1 and Experiment Unit 2 which are consistent in the following respects:

1: All the target phonemes occur word-final and they all occur in the fifth/seventh syllable from the beginning of the sentence and the second syllable from the end of the sentence for Experiment Unit 1 and the third syllable for Unit 2.
2: All the carrier words are monosyllabic.
3: The phoneme following each target phoneme is consistently /ð/: there for Experiment Unit 1 and the for Experiment Unit 2.

Eg. She cooked and ate the peas there. (target phoneme: /z/)
She tried to keep the peace there. (target phoneme: /s/)

Experiment Unit 1: Six short sentences in which three pairs of words which differed in voicing feature of word-final consonant were embedded (CAP-CAB, LEAK-LEAGUE, PEACE-PEAS) were spoken at normal speed by a male native English speaker and recorded on audiotape. The sentences were all grammatically and semantically appropriate and had a biased context for the carrier word and consequently the target phoneme. The carrier words were all monosyllabic nouns with the definite article the and the adverb there subsequently. For the purpose of the judgement phase, I created sentences in which the target phoneme occurs only once. The sentences were then digitized and stored on a file of a speech analyzing software called SoundEdit. By using this sound-wave editor, the second version of stimuli were extracted from the original sentences. The extracted items did not contain an appropriate sentential context so that subjects might be garden-pathed phonologically. For example, an item
the peas there was extracted from the sentence context She cooked and ate the peas there, and its pairing item the peace there was from the context She tried hard to keep the peace there. When recording, the speaker was asked to pronounce the target phonemes as they are realized in fluent speech so that voiced phonemes tended to be heard as voiceless.

Once this extraction process was properly done, then the six extracted items were randomized with the four items in Experiment Unit 2 and were recorded on audiotape again to use in the experiment phase 1. Likewise, the six original sentence contexts containing the extracted items were recorded on audiotape in a random order together with the four sentences in Experiment Unit 2 to use in the experiment phase 2. In randomizing the sentences, nine distracter sentences in which the target phonemes occur word-initially were inserted for fear that subjects would always pay attention to word-final phonemes and would detect the purpose of the experiment. The results of the experiment phase 1 show, at the same time, the extent to which these phonemes were hard to differentiate.

Experiment Unit 2: The fundamental procedure to produce the material follows the Experiment Unit 1. Four short sentences in each of which the past participle of a monosyllabic regular verb was embedded (GRABBED, HUGGED, PACKED, PASSED) were constructed. Each sentence has a biased context for the carrier word and consequently the target phoneme /t/ or /d/. For example, an item packed the bag was extracted from the sentence context She should have packed the bag. When recording, the speaker was asked to pronounce the target phonemes as they are realized in fluent speech so that they were likely to be deleted or considerably difficult to recognize. The results of the experiment phase 1 show, at the same time, that these phonemes were hard to recognize.

2.2 Procedure

There are two judgement phases. In the first phase, subjects heard 10 extracted items in a random order. An answer sheet was provided in which a judgement question was written in each single line (e.g., Did you hear /t/? YES / NO). As soon as they heard each trial, subjects were asked to uncover only one line and to respond to the question as quickly and intuitively as possible by circling either YES or NO. Ten trials were run successively with a seven-second pause between them.

In the second phase, subjects heard the 10 critical sentences mixed with nine distracter sentences (19 sentences altogether) in a random order. It is noted that for voiced consonants used in Unit 1, the question did not directly ask if the phoneme was
heard. For example, for identification of /z/, the question was “Did you hear /s/?”. This type of question is effective because the identification of these voiced phonemes needs strong rejection of hearing their voiceless counterparts. Thus the results of these questions should be reliable to a great extent. The rest of the procedure was the same as in the first phase.

Before proceeding with the task, the nature of the stimuli was fully explained aurally and visually to the subjects. Specifically, subjects were informed of the meaning of phonological notations such as /s/ and /z/ in order to avoid subjects’ simplistic matching of spelling and sound while doing the task. In both phases, stimuli were played using a portable cassette-CD player at comfortable volume in a quiet room, and subjects were seated at a distance of 50 cm from the loudspeaker.

It is noted that there seems to be no difference in restoration performance as a function of when the response is made. In Samuel’s (1981) study, it was reported that this kind of memory load appeared to play no role in phoneme judgement tasks. In my experiment, subjects’ responses to whether or not they actually heard the target phoneme were made after the completion of items or sentences. However, according to Samuel, their restorability (perception) cannot be impaired by such a delayed judgement. Thus subjects’ judgements in the present experiment are considered to be identical to their immediate perception of target phonemes.

2.3 Subjects

Twelve native English speakers who are graduate students of University of Cambridge participated in the experiment. They are neither linguistics nor psychology majors, and have no known hearing problem.

2.4 Results

For both Experiment Unit 1 (voiced/voiceless version) and Experiment Unit 2 (deletion version), each identification outcome was classified as a successful identification without preceding sentential context (Phase 1) and a successful identification with the context (Phase 2).

Table 1 presents a summary of successful identification rates of the target phonemes for Experiment Unit 1 and Experiment Unit 2, broken down by phoneme features.
**Table 1: Successful identification rates**

<table>
<thead>
<tr>
<th>Stimulus type</th>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Increments</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 (overall)</td>
<td>0.486</td>
<td>0.625</td>
<td>0.139</td>
<td>2.916**</td>
</tr>
<tr>
<td>voiced</td>
<td>0.306</td>
<td>0.583</td>
<td>0.277</td>
<td></td>
</tr>
<tr>
<td>voiceless</td>
<td>0.667</td>
<td>0.667</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Experiment 2 (overall)</td>
<td>0.208</td>
<td>0.646</td>
<td>0.438</td>
<td>5.326**</td>
</tr>
<tr>
<td>/d/</td>
<td>0.25</td>
<td>0.917</td>
<td>0.667</td>
<td></td>
</tr>
<tr>
<td>/t/</td>
<td>0.167</td>
<td>0.375</td>
<td>0.208</td>
<td></td>
</tr>
</tbody>
</table>

(** p < .01)**

**Table 2: Test Materials / Test Scores in Experiment Unit 1**

<table>
<thead>
<tr>
<th>Sentential Context</th>
<th>Token</th>
<th>Target phoneme</th>
<th>Phase 1 (Token Only)</th>
<th>Phase 2 (In Context)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Correct Identification Rate</td>
<td>N /12</td>
<td>Correct Identification Rate</td>
</tr>
<tr>
<td>1 She likes to wear...</td>
<td>the cap there.</td>
<td>/p/</td>
<td>0.833</td>
<td>10</td>
</tr>
<tr>
<td>2 She wanted to take...</td>
<td>the cab there.</td>
<td>/b/</td>
<td>0.166</td>
<td>2</td>
</tr>
<tr>
<td>3 She tried to stop...</td>
<td>the leak there.</td>
<td>/k/</td>
<td>0.666</td>
<td>8</td>
</tr>
<tr>
<td>4 She wants to join...</td>
<td>the league there.</td>
<td>/g/</td>
<td>0.333</td>
<td>4</td>
</tr>
<tr>
<td>5 She tried hard to keep...</td>
<td>the peace there.</td>
<td>/s/</td>
<td>0.5</td>
<td>6</td>
</tr>
<tr>
<td>6 She cooked and ate...</td>
<td>the peas there.</td>
<td>/z/</td>
<td>0.416</td>
<td>5</td>
</tr>
</tbody>
</table>

**Table 3: Test Materials / Test Scores in Experiment Unit 2**

<table>
<thead>
<tr>
<th>Sentential Context</th>
<th>Token</th>
<th>Target phoneme</th>
<th>Phase 1 (Token Only)</th>
<th>Phase 2 (In Context)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Correct Identification Rate</td>
<td>N /12</td>
<td>Correct Identification Rate</td>
</tr>
<tr>
<td>7 She might have...</td>
<td>grabbed the mouse.</td>
<td>/d/</td>
<td>0.333</td>
<td>4</td>
</tr>
<tr>
<td>8 She cannot have...</td>
<td>hugged the man.</td>
<td>/d/</td>
<td>0.166</td>
<td>2</td>
</tr>
<tr>
<td>9 She should have...</td>
<td>packed the bag.</td>
<td>/t/</td>
<td>0.083</td>
<td>1</td>
</tr>
<tr>
<td>10 She could have...</td>
<td>passed the exam.</td>
<td>/t/</td>
<td>0.25</td>
<td>3</td>
</tr>
</tbody>
</table>

**Experiment Unit 1: Identification of final voice opposition.**

Identifications in this category consist of four pairs of voice opposition in word-final (/p/-/b/, /k/-/g/, /s/-/z/). The data from the Phase 1 show how the sounds in the experiment were ambiguous in terms of voice opposition. As expected, word-final consonants tended to lose their voicing feature so that they were likely to be heard as voiceless. The results show the fact that other cues, such as the duration of preceding vowel or pre-articulating gestures for the final phonemes, were not evident in stimuli
used in the experiment. In fact, the correct identification rate of the voiceless series was considerably higher than that of the voiced series in Phase 1. The overall increment in the correct identification rate for the unit was .156. The sentence context effect was found marginally significant for overall tokens (t-stat = 2.916, df = 11, p = .007). Figure 1 presents graphically the successful identification rates obtained in the two phases.

Two major factors other than the acoustic cues apparently contributed to the result. First, sentential contexts used in the experiment might have been biased toward the appropriate phonemes, but appeared not to be strong enough to produce a significant improvement. Second, word frequency effect seemed to be influencing the utility of the sentential context in identification of words, and consequently, the critical phonemes. For example, as Table 4 shows, for cap-cab and peace-peas pairs, sentential context effect was relatively small for restoring the voiced phonemes; the increment rates for cab and peas were .166 and .250 respectively (the total number of occurrences in LOB Corpus for each pair was 24 to 9 and 159 to 8 respectively). In this case, if a word with a voiced target has a powerful opponent in terms of frequency, the restoration of the phoneme via sentential context may be less effective. In contrast, for the leak-league pair, greater context effect was found; the increment rate for league was .417 (the total number of occurrences in LOB Corpus for the pair was 3 to 47).

Thus, the data show that voiced target phonemes are more likely to be restored when the carrier word is more frequent than a carrier word containing its voiceless counterpart. In other words, sentential context effect will be enhanced with the help of such a positive frequency effect, whereas a negative frequency effect may block the feedback route from the higher-level context.

Figure 2 shows the results for the successful identification rates broken down by the two phoneme types, namely voiced and voiceless phonemes obtained in both phases.
Unfortunately, due to an insufficient number of sample sentences, we could not perform a statistical test for this phoneme-specific analysis. However, the figures presented here can lead us to predict that sentential context effect (in this particular case, semantic context) can influence the restoration of such word-final phonemes, especially devoiced voiced phonemes, although the context effect seems to be highly dependent on the degree of its semantic constraints, the intrinsic word frequency and the voicing features.

**Experiment Unit 2: Identification of morphemes for past participles.**

Identifications in this category consist of /d/ and /t/ as phonemes for past participles. In contrast to Experiment 1, the results from Experiment 2 show that sentential context was playing a greater role in restoration of the target phonemes. Table 1 shown earlier and Figure 3 below show the overall results.

The data from Phase 1 show that the critical sounds in the experiment were ambiguous across the two phoneme types, and that most subjects did not even recognize their existence. In other words, most of the subjects heard the past
participles as just verbs without any conjugation. This kind of garden-path effect was caused by deletion of stop consonants put between other consonants in fluent speech. However, the data from Phase 2 show that many of these missing sounds could be actually restored when they occurred in a predictable sentential context. In Unit 2, effects of all the sentential contexts were considered to be almost identical since all the carrier verbs were embedded in a modal-auxiliary-plus-\textit{have} structure such as \textit{might have grabbed}. Specifically, this type of context is considered to be more syntactic than semantic and must be more powerful than only semantically biased contexts. In fact, the sentential context effect in Unit 2 was highly significant (t-stat = 5.326, df = 11, \( p = 0.00012 \)) in a paired two-sample t-test.

![Figure 3: Overall successful identification rates](image)

When the data are broken down by the two types of phonemes, the results are even more dramatic. Figure 4 shows clearly the difference in restoration rates for /d/ tokens and /t/ tokens.

![Figure 4: Successful identification rates broken down by phoneme type](image)
As indicated in Figure 4, the restoration rate for /d/ tokens was much higher than that for /t/ tokens. This is presumably because the following consistent word, the definite article *the*, was always pronounced with actually voiced consonant /ð/, although in some other contexts, such as *What's the matter?*, ‘the’ may sometimes be pronounced as voiceless. Thus, a possible explanation is that the subjects’ impression of having heard /d/ was influenced by the quality of the following phoneme. Unfortunately, we cannot perform a statistical test for the phoneme specific analysis until more pairs of sentences are available. However, the results were appealing enough to predict that, at least in this particular constraint, sentential context plays a robust role in restoring the missing phoneme /d/.

3. DISCUSSION

The central question under consideration in the experiment was what role sentential context plays in the restoration of altered or deleted phonemes in fluent speech. The results provide a promising outlook for the direct role of sentential context effect, especially in the restoration of deleted morphemes. In Experiment 2, the data indicate that syntactic-sentential context plays the primary role in restoring the critical missing phonemes since most subjects failed to recognize the phonemes without the context. If it was true that subjects heard the missing phonemes in the contexts used in the test, this means that context is contributing directly to the perceptual processing. Subjects might be just hearing illusions, however if they were functioning as phonemes in their minds when processing the speech, then they were considered to be phonemes (Samuel, 1981; 1997).

As discussed earlier, lexical effect is a robust feedback process in phoneme restoration or perception. One plausible reason is that listeners use pre-stored lexical knowledge in processing speech (Connine, 1990; Connine and Clifton, 1987; Tanenhaus and Lucas, 1987). In this case, the pre-stored phonological form of a lexical item contributes directly to the computation of the auditory input being processed. Here, then, a question will arise: What kind of knowledge is pre-stored and what kind of representation is constructed during speech processing? If any knowledge used in speech processing is simply pre-stored or not, then the prevailing issue that the role of context is either true perceptual or post-perceptual will be preserved. However, if the knowledge has variability in the continuum of the two endpoints, namely pre-stored and constructed, then the role of context also will not necessarily be dichotomous.

In fact, not a few researchers claim that there are some constraints even in lexical effect (Cutler et al, 1987; Cutler and Norris, 1979; Fox, 1984). They argue that lexical...
effect is available only in the boundary region of speech continua between two canonical acoustic stimuli; only slow decisions allow the lexical route to win, and lexical effect then becomes active. These bottom-up theorists reject the direct role of lexical effect in phoneme processing. In addition to these two extreme views, in the Cohort Model of spoken word recognition (Marslen-Wilson, 1987; Marslen-Wilson and Tyler, 1980; Marslen-Wilson and Welsh, 1978), two phases in speech processing can be identified: a first bottom-up phase in which context effects are absent and a second interactive phase in which lexical effect can be detected. Thus, it seems to be a matter of degree whether context effect is a true feedback or a post-perceptual decision bias, namely, how well relationships among representations are established and pre-stored.

The above dispute over the role of lexical context seems to be parallel to our discussion of the role of sentential context. In other words, the nature of the role of sentential context may be dependent on how well the sequence of words is established and pre-stored. In this respect, the type of lexical items in the sequences of words used in the second unit of my experiment were syntactically constrained (e.g., *could have* plus *past participle*) so that it appeared that the knowledge of the sequence is more likely to be pre-stored and its processing more automatized. In fact, the results show a robust context effect. In contrast to these constraints, semantic context used in the first unit appeared to be less constrained so that the knowledge of a pre-stored set of phonemes was less available during the perceptual process, resulting in greater confirmation of bottom-up cues.

Apart from the degree of constraint of context, the methodology of tasks used for testing the role of context effect may vary the outcome to some degree. For example, the added/replaced task used in Samuel’s (1981) study tended to focus the listener’s attention on the speech sounds, emphasizing bottom-up processing. On the other hand, tasks such as shadowing (e.g., Marslen-Wilson and Welsh, 1978) tended to force the listener to rely more on top-down processing, absorbing their attentional capacity. Neither of these tasks is considered to be exactly replicating the way of processing speech sound in normal situation. In this respect, the task used in my experiment may reflect bottom-up processing to some degree because it was a kind of phoneme judgement task. Nevertheless, the results show the powerful role of sentential context effect. A drawback of the experiment is that we cannot monitor how well the subjects understood the nature of the task from the data collected. However, the methodology used in my experiment was able to reveal listeners’ true perceptual process to a great extent, since the question was to ask whether subjects actually perceived the critical phonemes or not. Thus, the present study is significant in that it has shown top-down context effect notwithstanding that the nature of the task was accentuating bottom-up
process.

Now let us return to our basic assumption that there is a level of phoneme representation in spoken word recognition and speech perception. This is an assertion of proponents for interactive models of speech perception such as TRACE. The results of my experiment support this view; the results of the second phase were an especially good example of this. Most subjects did not recognize the critical past participle morphemes without context, although a handful of the subjects were successful in recognizing them under the same condition. If there were no such phoneme level of representation, how could subjects restore the phonemes which they had missed in the absence of sentential context? The results suggest that the mechanism of human perception owes much to abstract representation. In other words, phonemic restoration plays a positive and definite role in actual perceptual processing of speech.

Some researchers argue against such an intermediate level of representation (phonemes) in spoken word recognition. In Klatt’s (1979) Lexical Access From Spectra model (LAFS), he proposes that phonemes may not be necessary for word recognition. He assumes that spoken words are recognized directly from an analysis of the (sufficiently invariant) input spectrum using a larger unit called diphones (phoneme-phoneme sequences). LAFS appears to be an elegant model of word recognition in that the model is capable of handling variability in spoken language such as speech rate, prosody and talker normalization. Furthermore, Stevens (1986) and Marslen-Wilson and Warren (1994) also stress the direct access of lexical entries on the basis of an acoustic analysis of the incoming speech signal without any intervention of phonemic representation. Instead of Klatt’s diphones, Stevens articulates the importance of linguistic binary phonetic features, whereas Marslen-Wilson and Warren highlight lexical representations specified in terms of phonologically distinctive features such as [+voice], [+nasal].

Apparently, however, all these bypass models fail to explain the results reported in the present experiment. Most subjects failed to recognize /d/ in a sequence like grabbed the mouse, even though the critical phoneme had both its left and right contexts. Also, most subjects heard /k/ in the league there, but many of them heard /g/ from exactly the same stimulus in the context She will join the league there. These results show how insufficient and error-prone the bypass models are. In these cases, phoneme level representation is crucial in word recognition and speech perception. Thus, the present research predicts a model of spoken-word recognition and speech perception such as TRACE in which there is an intermediate level of representation between lexical and feature level representations.
4. CONCLUSION

I showed through my experiment the possibility that sentential context can directly influence phoneme level processing. The results also provide evidence which supports the view that there is a level of phoneme representation in spoken word recognition. I suggest that the nature of the role of context may not necessarily be dichotomous between the purely true feedback and the post-perceptual decision bias; rather, it constitutes a continuum between the two ends, and is dependent on the degree of the constraint intrinsic to context.

References


Appendix I: Instructions for the procedure of the test

Informal Sound Judgment Test

This is an informal experiment which aims at finding out how native English speakers would respond to different speech sounds.

Please have a look at the following instructions of the procedure of the test, before proceeding with the task. Thank you.

Procedure:

Test 1
You will hear a small part of a sentence (which is both grammatically and semantically appropriate) which was extracted from a longer sentence by using a speech analyzer. As soon as you hear it, please uncover only one line of the Test Sheet. In the line you will find a question such as:

Did you hear /p/? YES / NO

Then, if you think you have heard the target sound (in this case /p/), then circle YES, and if not circle No as quickly and intuitively as possible.

Please note that the target sounds may occur in either word-initial or word-final.

Examples:
/z/ as in zoo or knees
/t/ as in take or heat, kicked
/d/ as in deep or mind, begged
/g/ as in good or pig, but not in college
/s/ as in sick or miss, but not in knees

You will continue to do the same task for 14 trials. Among the 14 trials you might hear completely the same recordings.

Test 2
You will hear a whole sentence (which is both grammatically and semantically appropriate) this time. The following procedure is the same as Test 1.

You will continue to do the same task for 21 trials.

Thank you for participating.

Experimenter
### Appendix II: Distracter Sentences used in the Test

<table>
<thead>
<tr>
<th>Distracter sentences used in the Test</th>
<th>Target Phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  She might have closed the door.</td>
<td>/d/</td>
</tr>
<tr>
<td>2  She should have stopped the car.</td>
<td>/t/</td>
</tr>
<tr>
<td>3  She would have bought the camera.</td>
<td>/b/</td>
</tr>
<tr>
<td>4  She must have got the train there.</td>
<td>/g/</td>
</tr>
<tr>
<td>5  She couldn't have put the book there.</td>
<td>/p/</td>
</tr>
<tr>
<td>6  She tried to cut the tape.</td>
<td>/k/</td>
</tr>
<tr>
<td>7  She might have seen the movie.</td>
<td>/s/</td>
</tr>
<tr>
<td>8  She loves the chair and the table there.</td>
<td>/t/</td>
</tr>
<tr>
<td>9  She is going to buy the doll there.</td>
<td>/d/</td>
</tr>
</tbody>
</table>