Effect of Lexical Status on Phonetic Categorization

Robert Allen Fox
Department of Linguistics, Ohio State University

To investigate the interaction in speech perception between lexical knowledge (in particular, whether a stimulus token makes a word or nonword) and phonetic categorization, sets of [bVC]–[dVC] place-of-articulation continua were constructed so that the endpoint tokens represented word–word, word–nonword, nonword–word, and nonword–nonword combinations. Experiment 1 demonstrated that ambiguous tokens were perceived in favor of the word token and supported the contention that lexical knowledge can affect the process of phonetic categorization. Experiment 2 utilized a reaction time procedure with the same stimuli and demonstrated that the effect of lexical status on phonetic categorization increased with response latency, suggesting that the lexical effect represents a perceptual process that is separate from and follows phonetic categorization. Experiment 3 utilized a different set of [b–d] continua to separate the effects of final consonant contrast and lexical status that were confounded in Experiments 1 and 2. Results demonstrated that both lexical status and contextual contrast separately affected the identification of the initial stop. Data from these three experiments support a perceptual model wherein phonetic categorization can operate separately from higher levels of analysis.

Understanding spoken language involves transforming the acoustic signal into an internal representation of the message conveyed by that signal. However, there is substantial disagreement among researchers as to how to model this perceptual process. One particular problem involves the hypothetical process responsible for converting the acoustic signal into an internalized sequence of individual phonetic segments—the process of phonetic categorization. There is disagreement over whether segments are (a) identified on the basis of the incoming acoustic information alone (as in a traditional bottom-up information-processing model, e.g., Fisoni & Sawusch, 1975); (b) identified only after the lexical item containing the segment has been identified (as in a strict top-down approach, e.g., Klatt, 1979, 1980b—a model that does not require the identification of individual phonetic segments at all); or (c) identified either "bottom-up" or "top-down" depending on which is the faster process (as in the dual-code hypothesis, cf. Foss & Blank, 1980). According to top-down views of phonetic processing, phonetic or phonological information is obtained from higher levels of processing (e.g., lexical access) rather than computed directly on the basis of bottom-up acoustic information.

Much of the disagreement in the phonetic literature stems from the fact that although there is probably a sufficient amount of acoustic information contained in the speech wave to allow identification of the phonetic segments under normal circumstances, because non-phonetic variables such as grammatical context can play a very significant role in segment identification, such bottom-up processing may be sufficient but not necessary. Examples of
top-down effects are legion, including the phoneme restoration effect (Warren, 1970, 1976), the intelligibility of grammatical versus ungrammatical speech in noise (Miller, Heise, & Lichten, 1957), and the fact that reaction times in phoneme monitoring tasks can be affected by context-dependent variables such as transitional probability (Morton & Long, 1976). A related issue is whether or not the phonetic categorization process operates as an autonomous level of analysis—that is, whether or not it makes its decisions on the basis of the available acoustic information and is not directly affected by higher levels of linguistic knowledge (as it is in an interactive perceptual model, cf. Marslen-Wilson & Welsh, 1978; McClelland & Rumelhart, 1981).

This article will examine a single aspect of these broad issues, namely, the extent to which the lexical status of a stimulus token (i.e., word–nonword) affects the phonetic categorization of its component segments and at what stage in the perceptual process this effect might occur.

It is clear that context may affect the categorization of phonetic segments in some biasing fashion. For example, Isenberg, Walker, Ryder, and Schweickert (1981) demonstrated that the phoneme boundary of a [bə]–[tʰə] ("the"–"to") continuum could be shifted significantly depending on the surrounding grammatical context. In particular, the perception of the initial consonant varied significantly in terms of whether the following word was a noun (which produced more the responses) or a verb (which produced more to responses). Similar effects have been reported by Garnes and Bond (1976) for a [b–d–g] place of articulation continuum and by Miller, Green, and Schermer (1982) for a [b–p] voice-onset time (VOT) continuum. However, there are studies that suggest that the lexical status of the stimulus token alone, unbiased by surrounding phonetic context, could produce significant variations in the perception of the token’s initial consonant. For example, Rubin, Turvey, and Van Gelder (1976) found that phoneme detection was faster when the target segment appeared in a word rather than a nonword. Such results are extremely suggestive either that lexical access occurs prior to segment identification or that lexical status biases the segment identification process in some fashion.

A study by Ganong (1980) was designed to address this issue directly. In particular, he demonstrated that the perception of ambiguous stop consonants from a VOT continuum could differ significantly depending on whether or not the token of which it was a portion made a word or nonword. For example, Ganong found that the [t–d] phoneme boundary differed significantly between a tash–dash continuum and a task–dask continuum. In each case the boundary was closer to the nonword endpoint, that is, more tokens were perceived as words than as nonwords.

Ganong noted that such results could be explained in several different ways. First, the lexical bias could be produced by a postcategorization correction process in which the candidate phonetic categorizations that did not make words would be changed to phonetic categorizations that did make words. Ganong termed this a categorical model. The correction would occur following category identification after which little or no acoustic information about the stop consonant would be available (reflecting the classic view of categorical perception model, cf. Repp, in press; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Ganong suggested a second model that allowed lexical information to affect and bias the process of phonetic categorization itself. This model, termed a criterion-shift model, would allow top-down (lexical) information to change the criterion against which VOT measures are compared for categorization. This criterion shift would occur at a point prior to or simultaneous with the phonetic categorization process itself. The criterion-shift model is thus interactive in the sense that it allows higher levels of linguistic knowledge to direct or determine lower level processing (cf. discussions in McClelland & Rumelhart, 1981, and Forster, 1979, concerning the issue of autonomy in levels of processing). Ganong argued that the criterion-shift model would concentrate the effect of lexical status at the phoneme boundary, whereas the postcategorization correction process would spread the lexical effect across the entire continuum. Because Ganong found that the lexical effect was much stronger at the phoneme boundary than at the endpoints, he argued that the data support an interactive top-down model.

Ganong’s specific conclusion was based solely on the interpretation of the shapes of
the identification functions and the locations of the interpolated phoneme boundaries. In particular, he made no attempt to examine the actual time course of the lexical effect. Ganong himself described at least one categorical model that would also explain the concentration of the lexical effect near the phoneme boundary. In this model, nonword categorizations would be reexamined but would not be corrected if the stimulus were far enough away from the phoneme boundary as to be consistent with only a single phonetic categorization. In other words, only relatively ambiguous tokens would be affected by the lexical correction process. This explanation runs into difficulty in terms of specifying how long unprocessed auditory information for the stop consonant can remain in an echoic store (cf. Crowder, 1971, 1978, 1982) but suggests that a categorical model could be constructed consistent with Ganong’s data. This issue will be reexamined at the end of this article in light of the data to be presented.

The present study was designed to (a) determine whether the lexical effect could be obtained with a [b–d] place-of-articulation continuum and (b) examine more directly the time course of the lexical status effect using a reaction time paradigm.

Experiment 1

Method

Stimuli. Four pairs of [b–d] continua were synthesized using the Klatt cascade/parallel synthesis program (Klatt, 1978, 1980a) implemented on The Ohio State University Linguistics Department’s PDP 11/23 computer. The four continua represented the initial consonants of syllables ending in [æ], [æb], [æd], or [æg]. The starting formant values for the eight different stimuli are specified in Table 1. The formant transitions were 30 ms in duration and went from their starting values to a formant pattern required for the vowel [æ] (F1 = 820 Hz, F2 = 1660 Hz, F3 = 2415 Hz). The first three formants were varied in this steplike manner, whereas the fourth and fifth formants remained fixed at 3400 and 3850 Hz, respectively. These transitions were slight modifications of those suggested by Klatt (1978, Appendix D). The vowel of each token was identical, but its formant structure changed slightly during its 300-ms duration, so that the first three formants ended at 650, 1490, and 2470 Hz, respectively. The fundamental frequency contour was falling throughout the token and, along with the changing quality of the vowel, produced a relatively natural-sounding speech token. The endpoint frequencies for the stimulus tokens with final consonants are shown in Table 2. The final consonant transitions were 35 ms in duration. Subjects (who did not participate in any of the three experiments to be described) identified the final consonants correctly 92% of the time in a forced-choice test. These syllable-final variations produced the word–nonword percepts shown in Table 3. The tokens that represent real English words have been indicated. The complete stimulus set includes two different word–nonword continua (lexically biased in opposite directions), one nonword–nonword continuum and one word–word continuum. Each continuum had eight members, thus producing 32 different stimulus tokens. A stimulus tape was constructed containing all stimuli in six different random orders with a 4-s interstimulus interval. A set of practice items preceded the test items on the stimulus tape.

Subjects. Fourteen subjects (paid volunteers who answered an advertisement in the college paper) participated in the experiments. All were native speakers of English with no known hearing impairment. All subjects were unfamiliar with the nature of the experiment.

Procedure. Subjects were required to indicate whether the initial segment of each stimulus was a [b] or a [d].

Table 1

<table>
<thead>
<tr>
<th>Stimulus token</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>[bae]</td>
<td>480</td>
<td>1401</td>
<td>2362</td>
</tr>
<tr>
<td>[bæb]</td>
<td>570</td>
<td>1349</td>
<td>2393</td>
</tr>
<tr>
<td>[bad]</td>
<td>461</td>
<td>1478</td>
<td>2424</td>
</tr>
<tr>
<td>[bed]</td>
<td>451</td>
<td>1516</td>
<td>2455</td>
</tr>
<tr>
<td>[bag]</td>
<td>442</td>
<td>1555</td>
<td>2486</td>
</tr>
<tr>
<td>[bag]</td>
<td>432</td>
<td>1593</td>
<td>2518</td>
</tr>
<tr>
<td>[dae]</td>
<td>422</td>
<td>1632</td>
<td>2549</td>
</tr>
<tr>
<td>[dæb]</td>
<td>413</td>
<td>1681</td>
<td>2617</td>
</tr>
</tbody>
</table>

Note. C = consonant; V = vowel.

Table 2

<table>
<thead>
<tr>
<th>Consonant</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>[b]</td>
<td>371</td>
<td>1150</td>
<td>2368</td>
</tr>
<tr>
<td>[d]</td>
<td>422</td>
<td>1832</td>
<td>2548</td>
</tr>
<tr>
<td>[g]</td>
<td>371</td>
<td>1924</td>
<td>2356</td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Stimulus token</th>
<th>CV</th>
<th>CVb</th>
<th>CVd</th>
<th>CVg</th>
</tr>
</thead>
<tbody>
<tr>
<td>[bæe]</td>
<td>[bæe]</td>
<td>[bed]</td>
<td>[bæg]</td>
<td></td>
</tr>
<tr>
<td>[dæb]</td>
<td>[dæb]</td>
<td>[dad]</td>
<td>[dæg]</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stimulus token</th>
<th>CV</th>
<th>CVb</th>
<th>CVd</th>
<th>CVg</th>
</tr>
</thead>
<tbody>
<tr>
<td>[bæe]</td>
<td>[bæe]</td>
<td>[bed]</td>
<td>[bæg]</td>
<td></td>
</tr>
<tr>
<td>[dæb]</td>
<td>[dæb]</td>
<td>[dad]</td>
<td>[dæg]</td>
<td></td>
</tr>
</tbody>
</table>
They were told only that they would be listening to speech tokens beginning with either of these two segments; they were told nothing concerning the lexical nature of the speech tokens to follow. Following data collection, two subjects were eliminated because their identification functions were flat, indicating they could not make reliable phonetic identifications.

Results and Discussion

The results of phonetic categorization, pooled across subjects, are shown in Figures 1 and 2. Figure 1 shows the identification function for the two word/nonword continua [bæg–dæg] versus [bæb–dæb]. The mean phoneme boundaries for each function appear in Table 4. The phoneme (or category) boundary represents the 50% cross-over point in the identification function. That is, it indicates the (interpolated) point on the stimulus continuum that would receive each phonetic categorization on half of the trials. There was a significant difference in the phoneme boundaries of the two functions with a shift in the boundary in the direction of the nonword token, \( t(11) = -3.70, p < .001 \), one-tailed; mean boundary difference = .88.\(^1\) This means that across both continua there tended to be relatively more word responses (i.e., the initial consonant was identified as a segment that would produce a real word). In addition it should be noted that, as in Ganong’s (1980) results, the lexical effect seems to be limited to ambiguous stimuli near the middle of the continuum and has little or no effect on stimuli near the endpoints.

The word–word and nonword–nonword identification functions are shown in Figure 2. There was a slight shift in the [baed–dæd] continuum toward the [d] endpoint relative to the [bae–dæ] continuum, \( t(11) = 2.18, p < .05 \), one-tailed, mean boundary difference = .21. This slight shift can also be interpreted in terms of the lexical status effect. Although the endpoints of the [bae–dæ] continuum might not be equally preferred subject responses, there is no obvious difference between them in terms of lexical status. However, in the [baed–dæd] continuum, although both endpoints represent real English words, [baed] (bad) is a more frequently occurring word than [dæd] (dad) (Carroll, Davies, & Richman, 1971). As has long been recognized (cf. Morton, 1979), word frequency has a significant effect on lexical access. If lexical status affects phonetic categorization (regardless of whether the effect occurs before, during, or following the categorization process), then significant differences (though perhaps small in comparison to the word–nonword differences) in the identification functions should be expected as a function of word frequency differences. After all, the word–nonword distinction itself can be viewed as a very large difference in terms of word frequency.

\(^1\) One-tailed significance tests are used where direction of shift is predictable.
Table 4

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Word percept</th>
<th>[ba–da]</th>
<th>[ba–dA]</th>
<th>[ba]–[dA]</th>
<th>[ba]–[dÅ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[ba–da]</td>
<td>4.48</td>
<td>4.47</td>
<td>4.78</td>
<td>5.05</td>
</tr>
<tr>
<td>2</td>
<td>[ba–dA]</td>
<td>4.45</td>
<td>4.46</td>
<td>4.37</td>
<td>4.56</td>
</tr>
<tr>
<td>Range 1</td>
<td>[ba–dA]</td>
<td>4.33</td>
<td>4.20</td>
<td>4.61</td>
<td>4.78</td>
</tr>
<tr>
<td>Range 2</td>
<td>[ba–dA]</td>
<td>4.30</td>
<td>3.97</td>
<td>5.45</td>
<td>4.95</td>
</tr>
<tr>
<td>Range 3</td>
<td></td>
<td>4.24</td>
<td>4.20</td>
<td>4.76</td>
<td>5.03</td>
</tr>
<tr>
<td>All RTs</td>
<td>[ba–dA]</td>
<td>3.87</td>
<td>4.39</td>
<td>3.81</td>
<td>4.21</td>
</tr>
</tbody>
</table>

Note. The phoneme boundaries for Ranges 1–3 were calculated on the basis of the supersubject data. RT = reaction time.

The results of Experiment 1 are thus consistent with Ganong’s data and, following his line of argumentation, again suggest that lexical status is having a direct effect upon the categorization process. However, his conclusions were based upon analysis of the shape of the identification functions themselves rather than upon information bearing directly upon the time course of the subjects’ identifications. In particular, we need to determine whether the influence of lexical status occurs simultaneously with the identification process (indicating a direct, top-down effect) or at some small (but measurable) time following identification (indicating a correcting or biasing effect upon the identification response to the already categorized segment).

Experiment 2

To address this question, a second experiment was conducted that used the same stimuli as Experiment 1 but required subjects to respond as quickly as possible. Subjects’ identifications as well as reaction times (RTs) were recorded, and the results were analyzed to determine whether the lexical effect increased, decreased, or remained stable as a function of delay between stimulus onset and subject response. The assumption that underlies this use of the reaction time paradigm is that if each level of perceptual analysis takes some processing time (however small), then the time it takes a subject to reach a perceptual decision may reflect the number of processing levels involved as well as the type of information used in the subjects’ decisions (see discussions in Pisoni & Tash, 1975; Posner & Mitchell, 1967). In terms of this particular experiment, if lexical status is directly influencing the categorization process and if such influence occurs either simultaneously with or prior to phonetic categorizations, then there should be a lexical effect even at the shortest latencies. One might then expect the lexical effect to remain stable at all latencies. This is based on the rationale that if the lexical effect is an integral part of the phonetic decision process itself or if it represents a process prior to phonetic categorization, then all subject responses, even those made immediately following categorization, have already been subject to the lexical status effect. However, if the lexical effect is due to a postcategorizational process, (i.e., a correction)—especially if it is under partially attention-driven, subject control—then one would expect the shorter latency responses to be less subject to the lexical effect than identifications with relatively long latencies. Obviously, this view requires that the postcategorization process be a separate level of analysis from phonetic categorization and that subjects respond before the correction process has been completed.

Method

Design. Same as in Experiment 1 except that the forced-choice task was timed.
Stimuli. Same as in Experiment 1.
Subjects. Subjects were obtained in the same way as in Experiment 1 but no subject was used in both experiments. Fifteen subjects participated in the experiment.

Procedure. The subjects were again required to identify the initial segment of each stimulus token as being either [b] or [d]. They did so by pressing one of two response buttons placed in front of them. The placement of the [b] and [d] buttons by the right or left hand was counterbalanced over subjects. Subjects listened to a set of practice tokens, and their responses and identifications were recorded by the computer. Before the test tape proper began, the experimenter discussed the practice results with the subject. In every case, the experimenter indicated to the subject that his or her reaction times were acceptable but should be improved during the actual test. This was done to bias the subjects toward speed rather than accuracy in order to avoid conscious, overt, corrections possibly related to the issue being addressed here. Following data collection, 2 subjects were eliminated because their overall identification functions were flat.

The stimulus tokens were played to subjects over high-quality headphones. In addition, the audio signal was simultaneously output to the analog-to-digital circuit of a PDP 11/23 computer. The computer noted the onset of energy in the A-to-D circuit and started an internal programmable clock. The computer next waited for the subject's key response and following the response noted the elapsed time between stimulus presentation and key depression as well as the identification response. These were stored on disk for later analysis.

Results and Discussion

In order to compare the results of Experiment 2 with those from Experiment 1, the data were first collapsed across all different reaction times. These aggregate identification functions, summed across all subjects, are shown in Figures 3 and 4. The mean phoneme boundaries for these data appear in Table 4. Note that these functions are very similar to those found in Experiment 1. As in Experiment 1, there is a significant difference in the phoneme boundaries of the two nonword-word continua, shown in Figure 3, with a relative shift in the direction of the nonword end-point, t(12) = 2.00, p < .05, one-tailed, mean boundary difference = .71. Comparison of the identification functions for the word-word and nonword-nonword continua, shown in Figure 4, again shows at least marginally significant differences in the interpolated phoneme boundaries with the [baed—aed] continuum shifted toward the less frequent endpoint relative to the [bae—aed] continuum, t(12) = 1.88, p < .05, one-tailed.

The results from Experiments 1 and 2 were compared using a 4 X 2 analysis of variance with the factors stimulus continuum (word–nonword, nonword–word, nonword–nonword, and word–word) and task (timed vs. untimed). Because there were unequal cell frequencies, a general linear models procedure was used (Ray, 1982). The results showed a statistically significant main effect of stimulus continuum, F(3, 92) = 4.46, p < .005, but the main effect of task was nonsignificant, F(1, 92) = 0.72, ns. The Continuum X Task interaction was also nonsignificant, F(3, 92) = 0.18, ns. Duncan's multiple range test showed a significant difference between the word–nonword and nonword–word continua at the .05 level.
The aggregate timed data thus showed the same lexical effect as obtained in Experiment 1, and there was no significant difference due to task, at least not in terms of the aggregate response data. However, if we more closely examine the identifications as a function of reaction times, quite a different pattern of subject responses is evident. For each of the four continua, the subjects' responses were divided into three different sets. In particular, identification functions were computed for responses from three separate ranges of reaction times: Range 1 (RT ≤ 500 ms), Range 2 (500 ms < RT ≤ 800 ms), and Range 3 (RT > 800 ms). These particular ranges were selected on the basis of the overall RT distribution. Each range accounted for at least 24% of all subject responses.

Figure 5 shows the identification functions (summed over all subjects) for categorizations of the [bæb–dæb] continuum (nonword–word) at these three different ranges of reaction times. A Wilcoxon matched-pairs test showed a significant difference (at the .05 level, one-tailed) in percentage of B responses between Ranges 1 and 2, and Ranges 1 and 3, but not between Ranges 2 and 3. As is evident from Figure 5, the phoneme boundary moves toward the nonword endpoint of the continuum as the latency between stimulus presentation and key response increases, indicating that subjects gave relatively more word responses as the delay between stimulus onset and key response increased. The phoneme boundaries for Ranges 1–3 are shown in Table 4.

Figure 6 shows the identification functions for the word–nonword continuum [bæg–dæg]. A Wilcoxon matched-pairs test again shows a significant difference in percentage of B responses between Range 1 and Range 2 (p < .05, one-tailed) and between Ranges 1 and 3 (p < .02, one-tailed), but no significant difference between Ranges 2 and 3. There were again relatively more word responses when subjects took longer to respond. The phoneme boundaries for Ranges 1–3 are shown in Table 4.

Figure 7 shows the appropriate identification functions to the nonword–nonword continuum [bæ–dæ]. Note here that the phoneme boundaries for Ranges 1–3 are shown in Table 4.

![Figure 5. Identification responses for [bæb–dæb] at three different reaction time ranges. (Range 1 = responses made at RT ≤ 500 ms; Range 2 = responses made at 500 < RT ≤ 800 ms; Range 3 = responses made at RT > 800 ms, where RT = reaction time for identification response.)](image-url)
Figure 6. Identification responses for [bɛg–dɛg] at three different reaction time ranges. (Range 1 = responses made at $RT \leq 500$ ms; Range 2 = responses made at $500 < RT \leq 800$ ms; Range 3 = responses made at $RT > 800$ ms, where $RT$ = reaction time for identification response.)

Figure 7. Identification responses for [bæ–dæ] at three different reaction time ranges. (Range 1 = responses made at $RT \leq 500$ ms; Range 2 = responses made at $500 < RT \leq 800$ ms; Range 3 = responses made at $RT > 800$ ms, where $RT$ = reaction time for identification response.)
boundaries, shown in Table 4, did not shift as a function of response latency. A Wilcoxon matched-pairs test showed no significant differences among the three RT ranges. Figure 8 shows the identification functions for the word-word continuum [bæd-dæd]. As in the nonword-nonword continuum, there were no significant differences among the identification functions (Wilcoxon matched-pairs test), although the phoneme boundaries, shown in Table 4, did shift in the direction of the less frequent word.

In order to analyze these data more completely, it would be desirable to calculate phoneme boundaries for each separate subject for each of the four stimulus continua at each of the three RT ranges and then analyze the data using the appropriate analysis of variance design. Unfortunately, given that there are only 5 subject responses for each of the 32 different stimuli, it was often the case that a subject responded only once or not at all to a particular token in a particular RT range. To overcome this problem, a variation of the jack-knife technique (Wainer, 1974) was used to create a set of 13 supersubjects. The response data for each supersubject were the summation of all but one of the 13 subjects’ responses in each of the three RT ranges. Each supersubject thus reflects data from 12 different subjects but a different combination of 12 subjects. This technique provided a sufficient number of responses for each stimulus token at the three RT ranges to warrant calculation of individual supersubject phoneme boundaries. It should be noted that the supersubjects so defined will underestimate the population variance in the subject responses. However, this will allow us to obtain at least an estimate whether or not there actually are significant shifts in subject responses as a function of latency.

These supersubject phoneme boundaries were analyzed in a 4 X 3 analysis of variance with the factors stimulus continuum and response latency. The results showed significant

Figure 8. Identification responses for [bæd-dæd] at three different reaction time ranges. (Range 1 = responses made at RT < 500 ms; Range 2 = responses made at 500 < RT ≤ 800 ms; Range 3 = responses made at RT > 800 ms, where RT = reaction time for identification response.)
main effects due to both stimulus continuum, 
$F(3, 144) = 52.8, p < .001$, and latency, $F(2, 
144) = 10.58, p < .001$, as well as significant 
Stimulus $\times$ Latency interaction, $F(6, 144) = 
25.57, p < .001$.

However, there is a major problem in un-
critically accepting a lexical status effect in-
terpretation of the results from Experiments 
1 and 2—a problem that arises from the fact 
that in both experiments lexical status is con-
founded with the identity of the final conso-
nant. In particular, it is possible that the ob-
tained boundary shifts are due to the influence 
of the final consonant on the perception of the 
initial consonant. Alfonso (1981) found this 
general pattern of contextual phonetic influ-
ence in a study utilizing stimuli similar to those 
used in Experiments 1 and 2 (although he ig-
nored lexical status as a possible factor). He 
presented subjects with $C_1VC_2$ ($C =$ conso-
nant; $V =$ vowel) monosyllables, where $C_1V 
represented one of the stimuli from either a 
The final segment $C_2$ represented a good ex-
emplar of [b, d, g, p, t, k].

The pattern of Alfonso's results were as fol-
lowes: Effects were contrastive, so that for the 
voiced continuum, the initial [b-d] boundary 
was shifted in favor of [d] by a final [b], and 
in favor of [b] by a final [d]. Final [g] also 
shifted the boundary in favor of [d], but by a 
much smaller amount than final [c]. The final 
voiceless consonants produced similar con-
trastive shifts. Thus the results of Experiments 
1 and 2 could be a product of phonetic contrast 
rather than of the lexical status effect. However, 
the contextual contrast hypothesis does not 
fully explain the boundary shift obtained here 
in favor of [b] when the final segment is [g]. 
In particular, Alfonso (1981) found a shift 
boundary shift in the opposite direction with 
final [g]. One possible alternative is that both 
lexical status and contextual contrast may sep-
arately influence the phonetic categorization 
process.

The combined effect of these two influences 
would correctly predict all boundary shifts 
found in Experiments 1 and 2 without ap-
ppealing to word frequency differences in the 
case of [bed-dæd]. However, attractive as this 
explanation may be, it must be tested by an 
experiment that attempts to unconfound these 
two variables.

Experiment 3

Experiment 3 was designed to control sep-
arately both the lexical status and phonetic 
context variables. In order to best compare 
the results with those obtained in Experiments 
1 and 2, the stimuli were again of the form 
$C_1VC_2$, where $C_1$ represented one of the con-
sonants from a [b-d] continuum, and $C_2$ was 
either absent or represented final consonants 
that made a word–nonword, nonword–word, 
or word–word continuum.

In order to control for the phonetic context 
effects, all $C_2$ segments represented alveolar 
consonants [t, t', z]—thus the contextual con-
trast effect should shift the [b-d] boundary 
uniformly in favor of [b]. Although the three 
alveolar consonants used ([t, t', z]) might have 
different degrees of effect on the phoneme 
boundary as a function of the different man-
ners of articulation, none of them would pre-
dict a shift in favor of [d]. The syllable-final 
variations produced the word–nonword per-
cepts shown in Table 5—the tokens that rep-
resent real English words have been indicated. 
As in Experiments 1 and 2, the complete stim-
ulus set includes four continua representing 
the following lexical variations: word–word, 
word–nonword, nonword–word, and non-
word–nonword. The particular phoneme 
boundary shifts predicted by the three hy-
potheses outlined earlier are shown in Ta-
ble 6.

Method

Stimuli. Four pairs of continua were synthesized using 
the Klatt synthesis program. The continua each represented 
a seven-step [b-d] place of articulation continuum and 
differed only in terms of their final segments. The four 
continua ended either in a [s], [st], [st'], or [st']. The 
formant transitions for each voiced stop were 30 ms in 
duration and went from their starting values (shown in 

---

2 Given the idiosyncratic nature of phonetic form and 
word status in English, it was impossible to find a set of 
stimuli of the form $bVc$ or $dVc$, where $V$ represent a lax 
vowel (necessary so that neither $bVb$ or $dVd$ represented a 
word) and $C$ represented bilabial or alveolar stop conso-
nants so that lexical status and contextual contrast would 
predict boundary shifts in the opposite directions (i.e., 
where the $bVb$ and $dVd$ tokens make a word and the $bVd$ 
and $dVb$ tokens made nonwords). Therefore, the final con-
sonants used in Experiment 3 are different from those in 
Experiments 1 and 2.
Table 5
Possible Stimulus Tokens and Word Percepts in Experiment 3

<table>
<thead>
<tr>
<th>Stimulus token</th>
<th>CVt</th>
<th>CVt</th>
<th>CVz</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ba] but</td>
<td>[bat] buzz</td>
<td>[bAZ] buzz</td>
<td></td>
</tr>
<tr>
<td>[da] dutch</td>
<td>[dat] buzz</td>
<td>[dAZ] buzz</td>
<td></td>
</tr>
</tbody>
</table>

Note. C = consonant; V = vowel.

Table 7) to a formant pattern for the vowel [a], whose first three formants were fixed at 620, 1220, and 2550 Hz, respectively. The initial transitions of the first three formants were varied in this steplike manner, whereas the fourth and fifth formants maintained values of 3400 and 3850 Hz, respectively, throughout. The transitions for the endpoint consonants were slight modifications of those suggested by Klatt (1978). The vowel of each token was identical. There was a falling fundamental frequency contour for each stimulus that produced a natural-sounding speech token. Three of the [ba-da] continua were synthesized with final consonants appropriate for the English segments [t], [tʃ], and [z]. Final [t] had a 40-ms transitional period where the first three formants went from those of the vowel [a] to 400, 1520, and 2600 Hz, respectively. The final [t] was synthesized with an aspirated release following a 40-ms period of silence by exciting the circuit with a hiss (aperiodic) source. Final [tʃ] was produced with similar formant transitions, but the fricative [ʃ] followed the 40-ms period of silence—synthesized by exciting the parallel resonance circuit with an aperiodic noise source using acoustic parameters similar to those suggested by Klatt (1978). Final [z] was produced with similar 40-ms formant transitions but, near the end of the transitional period, both voiced and aperiodic noise sources were allowed to excite the parallel circuit producing the appropriate acoustic output for the voiced English fricative [z]. The vowel for the [ba-daz] continuum was 80 ms longer than those for the [bat-dat] and [batʃ-datʃ] continua to mimic the vowel length distinction found in English. Again, the acoustic parameters utilized were slight modifications of those found in Klatt (1978). Subjects (who did not participate in any of the three experiments described) identified the final consonants correctly 99% of the time in a forced-choice test. Each continuum had 7 members, thus producing 28 different tokens. A stimulus tape was constructed that contained all stimuli in six different random orders with a 4-s interstimulus interval. A set of practice items preceded the test items on the stimulus tape.

Subjects. Eighteen new subjects (volunteers meeting a course requirement) participated in the experiment. All were native speakers of English with no known hearing impairments. All subjects were unfamiliar with the nature of the experiments.

Procedure. Same as in Experiment 1. Following data collection, 1 subject was eliminated because she could not make reliable phonetic identification.

Results and Discussion

The results of phonetic categorization, pooled across subjects are shown in Figures 9 and 10. The mean phoneme boundaries for each function appear in Table 4.

As is evident from the pattern of results, the location of the phoneme boundary seems to be affected by both lexical status and contextual contrast. In particular, there is a significant difference between the boundary of the [batʃ-datʃ] continuum and both the [ba-daz], $t(16) = 3.09$, $p < .007$, two-tailed, and [bat-dat] continua, $t(16) = 3.99$, $p < .001$, two-tailed. There is no significant difference between the [ba-daz] and [bat-dat] continua, $t(16) = 1.04$, $p > .30$, although clearly final [t] tends to shift the phoneme boundary slightly more in favor of [b] than does final [z]. There is a significant difference between the [ba-da] baseline continuum boundary and both the [bat-dat], $t(16) = 2.90$, $p < .01$, two-tailed, and the [ba-daz] continua, $t(16) = 2.67$, $p < .02$, two-tailed. If phonetic context alone were responsible for the phoneme boundary shifts, then we would expect the [batʃ-datʃ] boundary to be near that of [bat-dat], but this was not the case. There is no significant difference be-
Table 7
Starting Formant (F) Frequencies (in Hz) for
Initial Consonant Transitions in Synthesis
of [ba–da] Continua

<table>
<thead>
<tr>
<th>Stimulus item</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>484</td>
<td>1149</td>
<td>2418</td>
</tr>
<tr>
<td>2</td>
<td>472</td>
<td>1202</td>
<td>2444</td>
</tr>
<tr>
<td>3</td>
<td>460</td>
<td>1255</td>
<td>2470</td>
</tr>
<tr>
<td>4</td>
<td>448</td>
<td>1308</td>
<td>2496</td>
</tr>
<tr>
<td>5</td>
<td>436</td>
<td>1361</td>
<td>2522</td>
</tr>
<tr>
<td>6</td>
<td>424</td>
<td>1414</td>
<td>2548</td>
</tr>
<tr>
<td>7</td>
<td>412</td>
<td>1467</td>
<td>2574</td>
</tr>
</tbody>
</table>

tween the [ba–da] and [bat–dat] continua, \( t(16) = .66, p > .50 \). This suggests that the two effects are of equal strength, but in opposite directions. We interpret these data as supporting the contention that both lexical status and contextual contrast can affect the phonetic categorization process separately.

General Discussion

The results from this study demonstrate that (a) lexical status can affect the perception of a place-of-articulation continuum as it has been shown to affect perception of VOT continuum (Ganong, 1980) and the perception of tone (Fox & Unkefer, in press); (b) the effect is much stronger for acoustically ambiguous tokens (i.e., those near the phoneme boundary); (c) the degree to which lexical status affects a subject’s perception varies as a function of the latency between stimulus presentation and subject’s key response; and (d) both contextual contrast and lexical status may separately affect the phonetic categorization process. The results of Experiment 2 demonstrate that at relatively short latencies subject’s identifications are not significantly affected by whether or not the stimulus token represented a real English word.

There are several different (though interrelated) theoretical issues involved in interpreting these data with regard to effect of lexical status on phonetic categorization. These issues include determining the level of processing at which the lexical effect takes place and ascertaining whether a postulated phonetic level of processing operates autonomously with regard to higher levels of linguistic knowledge.

Given the limited scope of our study, the question of where the lexical status effect occurs will be limited to ascertaining whether it occurs prior to, simultaneously with, or following the phonetic categorization process. We will assume (without argument) that any hy-
A hypothetical lexical status effect must take place following lexical access; therefore, the question becomes one of determining when lexical access occurs relative to phonetic categorization.

One possible hypothesis is that lexical access always occurs prior to phonetic categorization. In this article we have assumed, at least implicitly, that a level of bottom-up phonetic processing exists. However, models of speech perception can be constructed that do not include phonetic processing as a separate stage of analysis. For example, Klatt (1979, 1980a) proposed a perceptual model in which words are identified on the basis of spectral shapes without requiring a level of analysis in which the individual phonetic segments are identified. One could view Morton's (1969, 1979) logogen theory in a similar light. Some speech researchers have gone so far as to suggest that "phonemes . . . are without direct perceptual basis . . . [and] have no direct relevance to perceptual processes leading to the comprehension of speech" (Warren, 1976, p. 409). Such models can be described as strict top-down views of phonetic processing.

The results presented here do not support a strict top-down approach. In particular, a top-down model predicts that lexical effects will appear in a subject's responses no matter how quickly he or she reacts. However, the identifications made at Range 1 in Experiment 2 show no such lexical effects. These data are consistent with those of Foss and Blank (1980, Experiments 1 & 2), which suggest that individual phonetic segments can be computed directly and do not necessarily result from recognition of a word (or other "higher" level unit of linguistic analysis). Even stronger evidence against the strict top-down approach has been obtained recently by Foss and Gernsbacher (1983). They argue that previous work (e.g., Morton & Long, 1976; Rubin et al., 1976) purporting to demonstrate that lexical status and/or semantic predictability produces responses based on a postlexical (phonological) code are best understood in terms of phonetic factors. Thus, we can reject the notion that lexical access always occurs prior to phonetic processing and, by implication, that the lexical status effect always occurs prior to phonetic categorization.

Rival models of speech perception maintain that phonetic processing is separate from lexical access (e.g., Foss & Blank, 1980; Foss, Harwood, & Blank, 1980; Newman & Dell, 1980; Pisoni & Tash, 1974) and that phonetic categorization occurs prior to or simultaneously with word recognition (the two processes operating in parallel). In such models the relative timing of phonetic identification and lexical access may differ potentially. For example, when the phonetic code is used to access lexical entries (e.g., in Foss & Blank, 1980), then phonetic identification precedes word recognition. However, lexical access may be speeded when preceded by a semantically related word in a sentence (Blank & Foss, 1978). This suggests that a phonological code (derived from the lexical entry accessed) may be available while the phonetic code is being developed, especially when contextual factors facilitate lexical retrieval (but see Foss & Gernsbacher, 1983, for an alternative hypothesis). The work by Marslen-Wilson and Welsh (1978) suggests, in particular, that lexical access may occur at least before all of the phonetic information of a word has been processed.

The Range 1 results of Experiment 2 demonstrated that subjects can respond on the basis of a phonetic code alone, unbiased by lexical status information. However, the results from Ranges 2 and 3 indicate that at longer latencies subjects' identification responses may be biased by lexical information. If both phonetic categorization and lexical access (i.e., word recognition) are separate processes, then each may have a different range of normally distributed completion times. The Range 1 results could reflect the fact that the mean time for phonetic categorization is shorter than for lexical access and that the responses correspond to decisions made on a purely phonetic basis. At Ranges 2 and 3, lexical access may have occurred before the identification response was made, and the phonological information obtained via word recognition could have "biased" the subject's decision.

The biasing could take several different forms, none of which may be eliminated on the basis of the results presented here. For example, if subjects have both a phonological code (derived via word recognition) and a phonetic code, they may normally respond on the basis of the phonological code alone. This would be consistent with the dual-code hy-
pothesis (Foss & Blank, 1980). Ambiguous stimuli in the word–nonword continua would access the lexical item corresponding to the word endpoint, allowing subjects to respond on the basis of the obtained phonological code. The unambiguous nonword stimuli would not access that word and could be identified on the basis of the phonetic code. Others versions of this hypothesis are possible.

A second model could represent the lexical effect as a correction process following phonetic categorization. Ganong (1980) discussed such a model but dismissed it on the basis that no acoustic information about the (possibly ambiguous) stop consonant is available following phonetic categorization (consistent with the view of Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; see Repp, in press, for detailed discussion of the categorical perception issue). However, there is some evidence that such information about stops may be present after categorization, although the memory trace may decay very quickly (e.g., Eimas & Miller, 1975, cited in Repp, in press; Pisoni & Tash, 1974). If such acoustic information is retained in an auditory memory store for some short amount of time, then a categorical model with a postcategorization correction process is a viable alternative to Ganong’s criterion shift model. That is, when the output of the phonetic categorization process is not a word (i.e., when the phonological code and the phonetic code differ) and the stimulus item is relatively ambiguous, then the response would be “corrected.”

Ganong (1980) argued that the lexical status effect occurred as a function of the direct interaction between lexical knowledge and the process of phonetic categorization. The results presented here as well as those discussed in Foss & Gernsbacher (1983) demonstrate that phonetic categorizations can be made without being affected by higher levels of linguistic processing. But on the basis of our data, it is premature to claim that the phonetic categorization process is strictly autonomous until we can better understand the precise mechanisms underlying the lexical biasing effect.

---

3 Foss and Gernsbacher (1983) conjecture that such nonword responses could be also made on the basis of a representation of the phonological “spelling” of the stimulus constructed from the phonetic code.

References


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

Received March 30, 1983
Revision received March 22, 1984