Phonological processes and the perception of phonotactically illegal consonant clusters

MARK A. PITT
Ohio State University, Columbus, Ohio

The perception of consonant clusters that are phonotactically illegal word initially in English (e.g., /tl/, /sr/) was investigated to determine whether listeners' phonological knowledge of the language influences speech processing. Experiment 1 examined whether the phonotactic context effect (Massaro & Cohen, 1983), a bias toward hearing illegal sequences (e.g., /tl/) as legal (e.g., /tr/), is more likely due to knowledge of the legal phone combinations in English or to a frequency effect. In Experiment 2, Experiment 1 was repeated with the clusters occurring word medially to assess whether phonotactic rules of syllabification modulate the phonotactic effect. Experiment 3 examined whether vowel open-thesis, another phonological process, might also affect listeners' perception of illegal sequences as legal by biasing them to hear a vowel between the consonants of the cluster (e.g., /tae/). Results suggest that knowledge of the phonotactically permissible sequences in English can affect phoneme processing in multiple ways.

An overarching goal of research in psycholinguistics is to delineate the structure and flow of information through the language processing system. A popular method of addressing this issue is to investigate how different types of linguistic information interact during processing. Interaction has been explored in areas ranging from the integration of semantic and syntactic information in sentence and discourse processing (Altman, Garnham, & Dennis, 1992; Boland, 1997; Boland & Cutler, 1996; MacDonald, Pearlmutter, & Seidenberg, 1994) to local contextual influences in phoneme processing.

Research on contextual effects in phoneme perception has been quite varied, including demonstrations of visual and auditory cue integration in phoneme identification (McGurk & MacDonald, 1976; Massaro, 1987), perceptual compensation in phoneme processing as a result of coarticulation differences from the preceding segment (Mann & Repp, 1981; Repp & Mann, 1981), and lexical and phonological influences on phoneme identification (Connine, Blasko, & Wang, 1994; Massaro & Cohen, 1983; Pitt & Samuel, 1995; Samuel, 1996). The present investigation focused on the last of these.

A number of researchers have argued that the phonology of a language is a rich source of information that could be exploited to facilitate auditory word recognition (Church, 1987a, 1987b; Frauenfelder & Lahiri, 1989; Frazier, 1987; Gaskell, Hare, & Marslen-Wilson, 1995). The case that is made in favor of this position is that much of the phonological variability in speech production is lawful and can be understood by consulting the phonology of the language. For example, vowels in unstressed syllables (e.g., schwa) can be deleted in spoken words, sometimes creating strings with phonotactically illegal sequences (e.g., tomorrow → tomorrow). Knowledge of English phonotactics could be used to recognize that /tm/ is not a permissible word-initial sequence. The intended vowel could then be recovered, leading to successful recognition of the word. Thus, phonological knowledge, like other forms of linguistic information (e.g., lexical) could aid word processing.

Increasing experimental evidence suggests that phonological knowledge is indeed used during recognition. Lahiri and Marslen-Wilson (1991) showed that listeners' responses in a gating task were predictable on the basis of the role a phonetic cue serves in the language. Vowel nasalization is nondistinctive in vowels in English. When such a vowel is heard without the next phoneme, English listeners will guess that the following consonant is a nasal. In Bengali, vowel nasalization is distinctive, making the identity of the following consonant (e.g., nasal or stop) unpredictable. Gaskell and Marslen-Wilson (1996; see also Gaskell et al., 1995) have reported related work in which the recognition system appears to compensate for place assimilation in speech production by applying knowledge of English phonology. Reaction times in an auditory–visual repetition priming task were equally fast when the prime was pronounced correctly (e.g., lean in lean bacon) or with the place of articulation of the nasal the same as the onset of the following word (e.g., lean in lean bacon). These results suggest that the lexical representation of lean was activated equally well by both primes. Importantly, there was a slowdown (21 msec) in response time when the mispronounced prime was presented in an unviable context, in which the nasal did not assimilate to
the place of articulation of the following stop (e.g., *team* game). Gaskell and Marslen-Wilson entertained the idea that phonological processes "inferred" the intended place of articulation when assimilation was viable (e.g., "tear bacon").

Findings from other studies suggest that a listener's knowledge of the phonotactics of English (permissible phoneme combinations) influence speech perception. Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk (1993; Jusczyk, Luce, & Charles-Luce, 1994; see Jusczyk, 1995, for a review) found that infants developed a preference for phoneme sequences in their native language at 9 months of age, but showed no such preference at 6 months. Cutting (1975; Cutting & Day, 1975) noted that phonotactically illegal sequences were never reported by listeners in phonological fusion experiments, in which a pair of words is presented dichotically and listeners must report what was heard. For example, *pay* was presented to the left ear and *lay* to the right. When the percepts fused, they were heard as *play*, never as *lpay*, which suggests that listeners' knowledge of English phonotactics influenced how the words blended.

Of particular relevance to the present study are data reported by Massaro and Cohen (1983; see also Brown & Hildum, 1956; Flege & Wang, 1989). Listeners had to categorize steps along a /r/-/l/ continuum as either /r/ or /l/. Each step was embedded in a consonant (obstruent) context so that the liquid formed the second consonant of a two-consonant cluster (e.g., /tr/). In the conditions of interest, the context phonemes were chosen so that the cluster formed at one endpoint of the continuum was phonotactically legal at the beginnings of words (e.g., /dr/, /sl/, /tr/) and at the other endpoint it was illegal (e.g., /dl/, /sr/, /tl/).

Listeners' classification responses showed a bias in favor of legal sequences. Steps at the /l/ endpoint were identified more frequently as /r/ when preceded by /t/ but never as /r/ when preceded by /s/. Just the reverse was found at the /r/ end of the continuum: Steps tended to be heard as /l/ when preceded by /s/, but as /r/ when preceded by /t/. Curiously, the /dl/ context, which might have been expected to produce an /rt/ bias equivalent to that of /tr/, showed minimal effects of phonotactic legality (Massaro & Cohen, 1983; Experiment 3).

Because labeling in the /s/ and /l/ contexts was always toward a legal sequence, one interpretation of these findings is that knowledge of the phonotactic constraints of English (i.e., rules of permissible phoneme combinations) affected processing of the liquid. An equally plausible (though not mutually exclusive) interpretation is that the results were due to the frequency with which the cluster occurs in the language. This can be thought of as an extreme version of a frequency effect, in which illegal clusters never occur word initially in English.

Although the first interpretation invokes a linguistically based mechanism to explain the outcome, no such commitment is required of the latter interpretation. The context effect might be no more than another demonstration of a general frequency effect, not the reflection of a process specific to language. Furthermore, even if the effect is specific to language processing, phonological processing per se is not necessary to produce the phonotactic context effect. The TRACE model of word recognition can simulate the phenomenon without relying on explicitly stored knowledge about English phonology (McClelland & Elman, 1986; see also McClelland, 1991; Massaro, 1989). The bias toward perceiving legal sequences comes about because there are more lexical entries with legal than illegal sequences. Because all lexical entries are activated when they match the speech input, there will be more top-down activation of phonemes that result in legal sequences (e.g., /t/ given a preceding /l/) than illegal ones (e.g., /l/ given a preceding /t/), biasing perception of legal clusters (e.g., /tr/). By being sensitive to the frequency with which clusters occur in the language, TRACE mimics what looks like the operation of a phonotactic rule.

The present investigation explored the perception of phonotactically illegal sequences with the aim of developing a deeper understanding of phonological influences in speech processing. In addition to addressing the cause of the phonotactic context effect, the influence of two other forms of phonological information on the perception of illegal clusters was investigated—rules of syllabification (Experiment 2) and vowel epenthesis (Experiment 3).

**EXPERIMENT 1**

In a frequency-based explanation of the phonotactic context effect, the size of the effect should correlate positively with the difference between the frequencies of pairs of clusters (e.g., /tr/ vs. /tl/). As the size of the difference increases, so should the context effect. Furthermore, this should hold true for any pair of clusters, not just those that are illegal. A strictly rule-based phonological account might not show sensitivity to variation in the frequency of a cluster, but only to whether the cluster is legal. In this case, phonotactic effects should be found only for illegal clusters and be similar in magnitude.

When Massaro and Cohen's (1983) data are examined with these predictions in mind, neither account is sufficient to explain the results. Although the /s/ and /l/ contexts yielded labeling shifts of similar size, supporting the claims of a phonological account, the /tl/ context produced a very small effect, a finding difficult to explain by a rule-based explanation.

To assess the accuracy of the frequency account, the frequencies with which the /dl/, /sl/, and /tr/ clusters occur in an on-line, phonologically transcribed version of Kučera and Francis (1967) were calculated and are shown on the left side of the top graph in Figure 1. The counts are independent of where in the word the cluster occurred (e.g., initially, medially) because a frequency explanation might not be sensitive to such information (e.g., the TRACE model). If the differences between the /l/
and /t/ counts for each context are compared, the predicted ordering of effect sizes in Massaro and Cohen's (1983) data should have been /t/ > /s/ > /d/., with /t/ being much larger than /s/, and /s/ being only slightly larger than /d/. Massaro and Cohen found /t/ = /s/ > /d/. Although /d/ was indeed smallest, the differences in effect size between contexts were not correctly predicted.

If the predictions of the frequency account are based on word-initial occurrences of clusters only (bottom graph in Figure 1), the frequency account fares even worse. The predicted ordering of effect sizes is /t/ > /d/ > /s/. Note that /s/ and /d/ are in the reverse ordering of what was found.

Experiment 1 was undertaken to explore further the viability of these two accounts. Liquid labeling was assessed across multiple contexts in which the two accounts make contrasting predictions. The three illegal contexts that were used by Massaro and Cohen (1983; Experiments 1 and 3) were combined with a /g/ context, which provided a means of dissociating phonotactic legality from cluster frequency. Both endpoints in the /g/ context (e.g., /gl/, /gr/) are legal, but the /t/ context is more frequent than the /l/ (see Figure 1). In fact, the magnitude of this difference (155, using the position-independent tallies) is larger than that for both the /d/ and /s/ contexts (66 and 72, respectively), so a large labeling bias should be present in the /g/ context. More precisely, a frequency account predicts that effect sizes should order as follows; /t/ > /g/ > /s/ > /d/). Predictions based on the word-initial tallies change only for /d/ and /s/, which swap places: /t/ > /g/ > /d/ > /s/). Because the word-initial and position-independent tallies yielded similar predictions, the frequency-based predictions will be discussed in terms of the position-independent tallies only.

A /br/ context served as a baseline from which effect sizes in the other contexts were measured. It was the only pair of stop–liquid clusters for which both clusters are similar in frequency. The slightly higher occurrence of /br/ might diminish the magnitude of phonotactic effects at the /l/ endpoint, but this should have an equal effect in all /t/-biased contexts.

The /g/ context also provides a useful test of the phonology account; /g/ was selected because it exhibits the largest frequency difference among stop–liquid clusters in which both liquids are legal continuations. Yet the phonology account predicts that no labeling bias should be found in this condition. The account again predicts phonotactic effects of similar magnitude in those contexts that have an endpoint that is phonotactically illegal (i.e., /d/, /s/, /t/).

**Method**

**Participants.** Thirty-one Ohio State University undergraduates, all native speakers of American English, participated in exchange for course credit. None reported hearing difficulties.

**Stimuli.** The Klatt (1980) synthesizer was used to create the obstruent–liquid–vowel continuum. The vowel /ɛ/ was chosen to avoid the creation of English words. The F1–F3 center frequencies were 711, 1743, and 2472 Hz, respectively.

For the context phoneme to influence liquid labeling, it was necessary to create a liquid continuum whose endpoints were not labeled as their respective categories 100% of the time while at the same time a clear and systematic change in labeling could be obtained across the continuum. Both spectral and temporal characteristics of the liquid were varied in continuum creation. An eight-step /t/–/l/ continuum was constructed using parameters listed in Massaro and Cohen (1983) and Samuel (1989) as guides. The primary parameter that varied was the F3 transition, which for the /t/ endpoint remained steady at 1752 Hz for 100 msec and then transi-
tioned to 2472 Hz over the next 100 msec. For the /l/ endpoint, F3 fell from 3173 Hz to 2472 Hz over the initial 250 msec. F1 and F2 followed a parallel trajectory at each endpoint. At the /r/ endpoint, the fundamental constant for 50 msec (474 Hz, 900 Hz) and then transitioned into the vowel formants over the next 100 msec. At the /l/ endpoint, they remained constant for 100 msec and then transitioned into the vowel formants over the next 50 msec. The six middle steps of the continuum were created by interpolating between these endpoint values in equal-sized steps. The fundamental frequency began at 160 Hz and dropped to 152 Hz by the end of the syllable. Voicing amplitude (AV) remained constant at 54 over the liquid and vowel, ramping to zero over the final 35 msec.

Steps on the /b/, /d/, /g/ and /l/ context continua were all 460 msec long. /b/ was synthesized with 50-msec formant transitions that began at 158 Hz for F1 and at 1100 Hz for F2. Voicing amplitude went from 50 to 60 over the first 10 msec before dropping to 54 over the next 10 msec. Frication amplitude (AF) decreased from 52 to 0 over the initial 5 msec. Bypass path amplitude (AB) increased from 0 to 53 during the initial 5 msec and then remained constant for 45 msec before dropping to zero over the next 5 msec. Formant transitions for /d/ were 55 msec long, with F1 and F2 starting frequencies of 211 and 1600 Hz. AV (54) began 15 msec after syllable onset. Frication increased from 20 to 50 over the first 10 msec and then dropped to 0 by 45 msec. /g/ formant transitions were 65 msec long, with F1 and F2 starting frequencies of 238 and 1732 Hz. AV increased from 0 to 58 over 5 msec beginning 15 msec after syllable onset. AF decreased from 60 to 0 over the first 20 msec. Formant transitions for /l/ were 65 msec long. F1 and F2 started at 378 and 1700 Hz. Voicing began 60 msec after syllable onset, rising from 0 to 54 in 5 msec. Frication amplitude was 60 at syllable onset and dropped to 0 by 50 msec. Aspiration amplitude was 30 at syllable onset, rose to 68 by 35 msec, and dropped to 0 by 65 msec.

Continuum steps in the /s/ context were 660 msec long. There were no formant transitions into the liquid. AV rose from 0 to 54 during the first 235 msec of the syllable. AF rose from 0 to 54 during the first 40 msec, remained constant for 150 msec, and then dropped to 0 over the next 40 msec. The amplitude of F6 was 64 for the first 20 msec, dropped to 60 for 145 msec, and then dropped to 0 over the next 50 msec.

The quality of the synthetic obstruents was assessed in a pilot experiment. Thirteen listeners categorized each obstruent embedded in the CCV context (legal continuum endpoints only) as one of five possible consonants 24 times. Fricative categorization was perfect. Stop categorization averaged 87% correct (range, 81%–93%), with 50% of the errors being made by 2 or 3 participants. Errors were not systematic across steps.

Exhaustive pilot testing with the full set of experimental materials was necessary to identify /t/ and /l/ endpoint tokens that were sufficiently perceptually ambiguous to show phonotactic influences. When the /r/-/l/ continuum was presented without a preceding obstruent in a liquid categorization task, the parameter values that were eventually selected yielded /r/ categorization responses of 89% for the /r/ endpoint and 26% for the /l/ endpoint (context effects failed to emerge with more extreme endpoint values). The difference in endpoint clarity (14%) might result in a slight underestimation of the size of the /s/ context effect on liquid labeling relative to the other context phonemes. As will be seen, the /s/ context effect was very robust, suggesting that the difference in endpoint clarity probably had minimal influence on the outcome of the experiment.

Equipment. Stimuli were synthesized at a 10-kHz sampling rate (12-bit resolution) and stored on hard disk. A microcomputer controlled stimulus presentation and response collection. Stimuli were low-pass filtered at 4.8 kHz before being amplified and presented to participants over headphones. Responses were collected using a four-button response box, with the left and right index and middle fingers pressing the buttons.

Procedure. Listeners were tested in groups of 4 or fewer, with each in a separate sound-attenuated cubicule. The experiment began with a familiarization session in which participants listened to examples of synthetic speech and to the endpoints of the continuum. In the test session, participants were instructed to classify the liquid in the syllable as belonging to one of four categories. Descriptions of the categories were printed above the four response buttons. Moving from one side of the response box to the other, the labels were “sure i”, “somewhat sure i”, “somewhat sure r”, and “sure r.” Four response choices were provided rather than two in an effort to increase the sensitivity of measuring differences between contexts. Each step on each of the five continua was presented 16 times for a total of 640 trials. Listeners were tested in two sessions that were separated by 2 days. In each, there were four test blocks of 80 randomly ordered trials, with each stimulus presented twice in a block. A 1.4-sec pause separated trials, and there was a 3-sec timeout after stimulus presentation. A rest break was provided at the end of Block 2. Twenty-four practice trials preceded the test session.

Results

Responses in the four categories were converted into numerical values, with “sure i” coded as 0 and “sure r” coded as 3. For each listener, a mean response score was calculated for each step on all continua. Listeners’ responses were then averaged and are plotted in Figure 2 as a function of the preceding context. Data in the /b/ context are represented with a dashed line. Contexts that were predicted by either account to yield an /r/ bias (/d/ vs. /g/ vs. /t/) are represented by filled symbols. The one context predicted to show an /l/ bias (/s/) is represented by an open symbol.

With the /b/ function as a reference, inspection of the figure shows that phonotactic context effects of differing magnitudes were found. At the /r/-/l/ endpoint, the /s/ function falls far below the /b/ function, indicating that the endpoint steps were heard far less often as /r/. Just the opposite pattern is present with the /t/ context at the /l/.

**Figure 2.** Mean /r/-/l/ labeling responses as a function of obstruent context in Experiment 1 (clusters occurred word initially).
Discussion

The results provide little support for a frequency-based account of the phonotactic context effect, which predicted an effect size ordering of /t/ > /g/ > /s/ > /d/. What was obtained was /s/ ≈ /t/ > /d/ > /g/. If the frequency account were correct, there should have been a strong positive correlation between the size of the phonotactic effect in each context and the size of the difference in /r/ and /l/ frequency counts in each context (Figure 1). Although positive, the correlation was quite weak (r = .15). Perhaps most damaging to the frequency account was the failure to find any hint of an effect in the /g/ context, where a large one was expected given the much more frequent occurrence in English of /gt/ than /gl/. To the extent that these data provide evidence against a frequency account, they are also damaging to models in which a frequency-sensitive mechanism is solely responsible for the effect (e.g., TRACE; see Gaskell & Marslen-Wilson, 1996, for a related discussion).

An explanation based on rules of phonotactic permissibility fares somewhat better. It successfully predicts that the phonotactic bias should be found only with illegal clusters (/dl/, /sr/, /tl/), not legal ones. It comes up short only in not accounting for the smaller effect found with /d/ than with /s/ and /t/. Some factor other than phonotactic permissibility may modulate phonological influences.

One possibility, which is also phonologically based, centers on a consideration of the memory representation of phonemes and how this might affect processing. Marslen-
other forms of phonological information might also affect the emergence of the phonotactic context effect. This idea was explored in Experiment 2 by testing whether similar labeling biases are obtained when clusters occur word medially.

Although clusters such as /tl/ and /dl/ do not form the onsets of words, they do occur in the middle of multisyllabic words (e.g., Atlantic, maudlin). Phonotactic illegality is avoided in these instances by syllabifying the words so that the initial consonant is placed in the coda of the first syllable and the liquid is placed in the onset of the following syllable (e.g., At-lan-tic, maud-lin). Treiman and Zukowski (1990; see also Treiman & Danis, 1988) showed that listeners apply this rule and others when syllabifying strings. For example, listeners virtually always syllabified words with medial clusters such as /tl/ and /dl/ between the stop and liquid, whereas legal clusters were frequently placed in the onset of the second syllable (e.g., a-pron, Ma-drid). Such outcomes are most readily explained by assuming that rules of English phonotactics are applied during syllabification.

One particularly intriguing aspect of these findings is that syllabification of words such as Atlantic is not what would be expected on the basis of an analysis of its acoustic realization. The consonant closure duration preceding burst release provides an obvious point at which to divide the word so that the illegal cluster is perceived in the onset of the second syllable (e.g., A-tlantic). The fact that the stop consonant can be heavily coarticulated with the liquid reinforces such an organization. Yet rules of syllabification are applied during processing that undo this organization and place the stop and the liquid in separate syllables.

During speech processing, does this powerful phonotactic rule of syllabification take precedence over that responsible for the phonotactic context effect? If so, the consonants of a word-medial illegal cluster would be perceived as belonging to different syllables before knowledge of phonotactic legality could affect liquid processing; the context effects observed in Experiment 1 should disappear or at least diminish in magnitude. That is, all labeling functions should overlap the /bl/ function across the entire continuum. If large labeling biases are still found, it would be likely that local (e.g., phonotactic) influences take priority in processing, with syllabification rules being applied later.

Method

Participants. Twenty-eight new listeners from the same population as Experiment 1 participated.

Stimuli. To create syllabic strings, the syllable /ma/ was synthesized and prepended to all of the stimuli of Experiment 1, except for the /gl/ context continuum, which was dropped from the design because it yielded no context effect. /ma/ was chosen because no English words were formed when combined with the CCVs of Experiment 1. /ma/ was 320 msec long with the formant frequencies of the vowel identical to those of the vowel in Experiment 1. Transitions for F2 and F3 were 80 msec, with F2 and F3 starting frequencies of 1243 and 2030 Hz. F1 began at 711 Hz, and from 80 msec to syllable offset it fell steadily to 692 Hz. F0 remained constant at 160 Hz. The nasal zero frequency began at 450 Hz at syllable onset, and at 80 msec it steadily dropped to 280 Hz by syllable offset.

The acoustic realization of a consonant can change depending on the context in which it is spoken. Of the four obstruents used, /t/ would likely undergo the most variation when produced in medial versus initial positions. Aspiration, a cue to identification, is present when /t/ occurs initially, but absent or greatly reduced when it occurs medially. /t/ was not resynthesized to take this change into account, however. Although failing to do so might have made the stimuli mildly artificial, I felt that it was more important to hold the stimuli constant across Experiments 1 and 2 so that the manipulation of interest could be assessed without there also being changes in stimulus characteristics.

Because other phonological characteristics of speech can affect syllabification (e.g., word stress, vowel tenseness; see Treiman & Danis, 1988), two pilot experiments were run to understand the stimulus conditions in which the labeling results were found. In the first, the endpoints of each continuum were presented to 13 listeners who had to repeat each syllable, but with the order of the syllables reversed (e.g., mabla → blama). Of interest was whether the illegal stop-liquid clusters would split apart (e.g., matla → lamat) or remain intact and form the onset of the first syllable (e.g., matla → tlam). Listeners showed a clear preference for placing the consonants in separate syllables than in the onset of the first syllable (66% vs. 34%, respectively). That phonological processes influence syllabification of the synthetic disyllables suggests that these stimuli are processed in a manner similar to that of naturally spoken utterances.

The results of a second pilot experiment, in which 16 listeners classified the continua endpoints as receiving primary stress on the first or second syllable, indicated that the disyllables were very nearly neutrally stressed. Responses averaged 57% (50% was perfect neutrality), a slight bias toward hearing the stimuli as being stressed on the second syllable.

Procedure. Pretesting indicated that a four-choice task did not yield data that were noticeably different from a two-choice task, so the latter was used. One response button was labeled "i," the other "I." The remaining methodological details were identical to those of Experiment 1.

Results and Discussion

Proportion of /t/ responses was calculated for each participant across all continuum. The aggregate listener data are shown in Figure 3 as a function of context. With the exception of the /d/ context, the data resemble those of Experiment 1. The /s/ context produced the smallest proportion of /t/ responses at the /t/ endpoint [22% effect, F(1.27) = 4.10, p < .05], and the /t/ context produced the largest proportion of /t/ responses at the /t/ endpoint [17% effect, F(1.27) = 2.95, p < .10]. In comparison with the corresponding contexts of Experiment 1, these two outcomes represent drops in effect size of 33% and 47%, respectively.

Although these data suggest that effects of phonotactic legality were weaker when the clusters occurred medially, other aspects of the results were unexpected and indicate that a pooled analysis of listeners' data obscured the true nature of phonotactic influences. To begin with, instead of the /d/ function rising slightly above the /b/ function, as was found in Experiment 1, it dropped below the /b/ function over most of the continuum (~7% reversal). Also, compared with Experiment 1, labeling functions were shallower and variability among listeners
was almost twice as great (Experiment 1, $SD = .17$; Experiment 2, $SD = .31$). The latter difference is a likely explanation for the statistically nonreliable phonotactic effect in the /t/ context.

Closer inspection of the labeling functions among participants revealed three categories of response patterns: Responses never fell below .5 across the continuum (/r/-dominant responders); responses never rose above .5 (/l/-dominant responders); responses crossed .5, ranging from at least .8 to .3 (/r/+ /l/ responders). When each listener's data were grouped by this classification, a clearer picture of phonotactic context effects emerged. The data are shown in the top part of Table 1. Listed in each cell is the number of listeners whose data fell into one of the three categories.  

In the /b/ context, 17 of 28 listeners heard the continuum change from /r/ to /l/. Six listeners heard it primarily as /r/ and five as /l/. A similar outcome was obtained in the /d/ context. Eighteen listeners produced full functions, and there was a slight /r/ bias for a majority of the remaining listeners.

Large phonotactic effects were obtained in the /s/ and /t/ contexts. These effects were present in the /r/-dominant and /l/-dominant functions. When the context was /t/, 15 listeners' data fell into the /r/-dominant category, whereas the data from only 2 fell into the /l/-dominant category. Just the reverse of this was obtained in the /s/ context, where there were 12 /l/-dominant listeners but just 4 /r/-dominant listeners. Chi-square tests were performed to assess whether the patterns of classification differed from chance. None reached significance in the /b/ and /d/ contexts. In the /s/ and /t/ contexts, reliable effects were obtained ($\chi^2 = 4.00, p < .05$; $\chi^2 = 9.94, p < .001$, respectively). The interaction of /r/-dominant and /l/-dominant data across the /s/ and /t/ contexts was also reliable ($\chi^2 = 13.94, p < .001$).

Even though a sizable phonotactic effect was obtained in the /s/ context, these data were noisier than in the other contexts. Seven listeners' functions were atypical and did not fall neatly into one of the three response categories (e.g., proportion /r/ was higher at the /l/ than at the /r/ endpoint and fluctuated erratically above and below the endpoint values in the middle steps). Because it seemed inappropriate to place the data from these ambiguous responders into one of the three categories, their data were categorized separately.

Although the data from these ambiguous responders do not compromise the results, it is useful to point out that they are an exception to the rule. The results from the /s/ and /t/ contexts in a similar experiment are shown in the bottom half of Table 1. The only methodological difference from Experiment 2 was that the initial nasal was spliced off of all stimuli to examine whether variation in the syllabic structure of the initial syllable (V instead of CV) affected labeling. As can be seen, the results tell the same story, but more cleanly. In the /s/ context, there were 15 more /l/-dominant than /r/-dominant listeners ($\chi^2 = 8.33, p < .001$). In the /t/ context, the opposite outcome was found, with 11 more /r/-dominant than /l/-dominant listeners ($\chi^2 = 7.12, p < .001$). The interaction across contexts was again reliable ($\chi^2 = 15.45, p < .001$).

Whether examined overall or looked at more closely as a function of listener response pattern, phonotactic context effects were found word medially. The subanalysis revealed that massive effects were obtained in the /s/ and /t/ contexts. Many listeners exhibited complete, not just partial, lifting and lowering of the functions at the illegal endpoints. Why larger effects were found medially than initially is likely due to the fact that liquid identification was more difficult word medially because of partial masking by the first syllable. The same reason, coupled with the fact that the liquid continuum had partially ambiguous endpoints, is also the most likely explanation for the wide variation in labeling among listeners in Experiment 2. In experiments with unambiguous continuum endpoints, listeners typically produce much more similar and complete labeling functions.

The data obtained in the /d/ context are the most equivocal; the only evidence of phonotactic influences is that two more listeners were classified as /r/-dominant than

### Table 1

Number of Listeners Whose Data Exhibited One of the Four Types of Labeling Functions in Each Phoneme Context

<table>
<thead>
<tr>
<th>Type of Labeling Function</th>
<th>/b/</th>
<th>/d/</th>
<th>/t/</th>
<th>/s/</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/r/-dominant</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>/l/-dominant</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>/r/ + /l/</td>
<td>17</td>
<td>18</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Ambiguous ( /s/ only)</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td><strong>Replication of Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/r/-dominant</td>
<td>14</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/l/-dominant</td>
<td>3</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>/r/ + /l/</td>
<td>13</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
/l/-dominant. Although the small effect is in keeping with what was found in Experiment 1, its virtual disappearance word medially might indicate that rules of syllabification had priority in processing. The robust phonotactic effects found in the /s/ and /t/ contexts provide strong evidence against this position, however, and suggest that rules of phonotactic permissibility were applied prior to rules of syllabification. Of course, it would be useful to assess the generality of the present results to disyllables that vary in ways that are known to affect syllabification (e.g., syllable of primary stress).

**EXPERIMENT 3**

The results of Experiments 1 and 2 suggest that knowledge of English phonotactics biases processing of liquids in illegal clusters so that they are heard as legal clusters. Another phonologically based method by which to make an illegal sequence legal is to insert a vowel between the two consonants (e.g., /tlae/ → /tælae/). This phenomenon, known as vowel epenthesis, can be heard in English if one attends closely to the pronunciation of the consonants of a cluster in exaggerated or slow speech. The initial consonant is often followed by a reduced vowel (e.g., schwa) to keep the utterance intact as a single word and can result in the perception of an additional syllable in the word (e.g., /grelt/ → /gærelt/).

Evidence suggesting that epenthesis is a phonological process operating during language processing can be found in loan words borrowed from other languages. For example, the Japanese language does not allow word-initial consonant clusters. English loan words with such clusters are modified to conform to Japanese phonology by inserting a vowel between the consonants of the cluster (e.g., /straij/ → /suтораикu/; see Takagi & Mann, 1994).

Is vowel epenthesis another method by which phonotactic processes affect the perception of illegal clusters such as /sr/ and /tl/? This question was addressed in the final experiment. The endpoint steps of the /t/ and /s/ context continua (/sr/, /sl/, /tr/, /tl/) were presented to listeners, who had to judge whether the utterances had one or two syllables. The steady-state portion of the liquid was progressively lengthened in each of the four stimuli to create the impression of schwa emerging between the obstruent and liquid.

If epenthesis is another means by which the perceptual system biases processing of illegal clusters, then two-syllable responses should occur more often with illegal clusters (e.g., /sr/, /tl/) than legal clusters (e.g., /sl/, /tr/). Put another way, illegal clusters might be more fragile than legal ones in terms of the consonants forming a unitary percept, and thus more susceptible to epenthesis. No differences should be found as a function of legality if phonotactic processes do not function in this manner.

**Stimuli.** /s/ and /t/ were used as the context phonemes because they produced the largest phonotactic effects in the preceding experiments. Two four-step continua were created with each context, one with /t/ and one with /l/. The synthesis parameters of the /sr/, /sl/, /tr/, and /tl/ endpoint tokens in Experiment 1 were used. The only change was that the F1–F3 steady-state portions that corresponded to the liquid were altered. For the two continua with /t/, the one-syllable endpoints had F1–F3 steady-state durations of 80 msec. For the two continua with /l/, the one-syllable endpoints had F1–F3 durations of 130 msec, and an F3 duration of 280 msec. Thirty msec were added to all of these values at each successive step, yielding a continuum whose endpoints differed by 90 msec. The range of values was selected based on piloting in which duration varied led to perceived changes in the number of syllables in the utterance. Syllable duration was 660 msec and remained constant across steps. The identity of the liquid was not noticeably affected by lengthening the steady-state portion of the vowel, in large part because the F3 formant transition, the primary cue to liquid identity, occurred after the steady-state portion. Perceptually, the steady-state portion sounded like /sr/.

**Procedure.** The procedure was similar to that of Experiment 1. Listeners were first introduced to synthetic speech and then instructed on the experimental task. Examples of one- and two-syllable words were provided to listeners to ensure that they understood which dimension to attend when categorizing the utterances as one or two syllables. No other instructions were provided on syllabification of the stimuli. One response button was labeled "one syllable," the other "two syllables." Two blocks of 160 randomly ordered trials were presented, with each step presented 10 times per block, for a total of 20 presentations per step. The pause between trials was 2 sec and there was a 3-sec timeout. The experiment was completed in one test session and there were eight practice trials.

**Results and Discussion**

Each listener's data were scored as the proportion of one-syllable responses to each step in each continuum. The averaged data are plotted in Figure 4 as a function of the four continua.

![Figure 4. Mean proportion one-syllable responses as a function of context and steady-state duration of F1–F3.](image-url)

**Method.**

**Participants.** Twenty-one new listeners from the same pool as the preceding experiments participated.
Overall, listeners tended to hear the stimuli as containing two syllables. Nevertheless, there was a bias toward labeling illegal clusters as being two syllables in length, and the bias held across the continuum. This outcome can be seen most easily by examining the data in each context separately. In the /t/ context (open symbols with unbroken lines), the /t/ function is always below the /tr/ function, indicating that utterances with the illegal cluster (/t/) were heard as containing two syllables more often than those with a legal cluster (/tr/). Just the reverse occurred in the /s/ context (filled symbols with dotted lines), where the /sr/ function is always below the /sl/ function. /sr/ was heard as being two syllables in length more than its legal counterpart /sl/. A two-way analysis of variance with context and liquid as factors revealed that the interaction of the two variables was reliable, indicating that one-syllable responses to legal sequences (/sl/, /tr/) were reliably greater than those to illegal sequences (/sr/, /t/) \( F(1,20) = 6.75, p < .02 \).

The pattern of labeling differed in the two contexts. In the /t/ context, the magnitude of the epenthesis effect varied minimally across the continuum (one-syllable responses to /t/ were .7-.12 higher than to /t/) and showed no systematic change in magnitude as formant duration increased. The interaction of liquid with formant duration was not reliable \( F(1,20) = 1.85, p < .19 \). In the /s/ context, the size of the effect increased across the continuum, with one-syllable responses to /s/ being .04 greater at the 0-msec endpoint and .22 greater at the 90+ msc endpoint. This change in effect size was reliable \( F(1,20) = 3.75, p < .02 \). The three-way interaction of context, liquid, and formant duration did not reach significance.

These results provide another demonstration of how phonotactic knowledge can affect the perception of illegal sequences: There is a bias toward perceiving a vowel between the consonants of an illegal cluster. Functionally, the epenthesis effect may be no different from the phonotactic context effect. In both phenomena, the context phoneme alters how the following section of speech is perceived. Which effect is found in a given situation might depend, among other things, on the clarity of the speech signal. If the liquid is perceptually ambiguous, context would likely bias processing in favor of the phonotactically legal sequence. If the liquid is a clear token, context might not as easily alter the identity of the liquid. Instead, the utterance could be reinterpreted as having an additional (epenthetic) vowel.

**GENERAL DISCUSSION**

The findings of this study provide further evidence that knowledge of English phonology affects speech processing. This conclusion is based on the results of Experiments 1 and 3, which, when taken together, demonstrated that phonotactic knowledge affects phoneme processing in two ways. In Experiment 3, the frequency of vowel epenthesis was shown to be affected by knowledge of English phonotactics. The labeling biases observed in Experiment 1 were better explained by an account based on listeners' knowledge of the permissible phoneme sequences in English than an account based on the frequency with which such sequences occur. The latter conclusion could be strengthened by generalizing the null result obtained with /g/ to a much less frequent context phoneme that also exhibits a large /t/ or /l/ bias and yields a continuum whose endpoints are phonotactically legal. Such a finding would demonstrate that the labeling bias is independent of the frequency of the context phoneme. Unless loan words are used, which pose other challenges, no such minimal pairs exist in English. It therefore seems prudent to use caution in generalizing the results beyond frequent context phonemes.

The present investigation also showed that syllabification processes had negligible effects on the phonotactic context effect (Experiment 2). Large labeling biases in favor of phonotactically legal sequences were obtained word medially, just as they were word initially.

When considered in the context of related work on phonological processing (Gaskell & Marslen-Wilson, 1996; Treiman, 1989; Treiman & Zukowski, 1990), an intriguing aspect of the present findings is that they suggest that the domain of influence (i.e., levels of representation) of phonotactic and syllabification processes are not the same. This claim is motivated by the failure to observe syllabification effects on liquid identification in Experiment 2, when such effects have been found repeatedly in other tasks (Treiman & Danis, 1988; Treiman, Gross, & Cwikiel-Clavin, 1992; Treiman & Zukowski, 1990). If all phonological knowledge affected all stages of processing, then at a minimum there should have been a noticeable reduction in the magnitude of the phonotactic effect. Yet the subanalysis showed that many listeners exhibited very robust labeling biases, hearing the entire liquid continuum mostly as /t/ or /l/ in illegal contexts.

As has been suggested for the application of rules of syllabification (Treiman & Zukowski, 1990), there might be an order of precedence in the application of different classes of phonological knowledge. For example, those rules that operate on a local scale (e.g., adjacent phonemes) would be applied before those that operate on a more global scale (e.g., whole word). Such an ordering could maximize processing efficiency and potentially minimize competition between them. Local rules would operate early and on small segments of speech. Further support for this idea comes from a recent study by Halle, Segui, Frauenfelder, and Meunier (1998), who examined the earliness with which phonotactic influences manifest themselves during perception. They used a gating task in which French listeners heard progressively longer segments of the initial portion of illegal clusters (/dl/ and /dl/) that formed the onsets of two-syllable pseudowords. Within the first 100–150 msec of cluster onset (Gates 4–5), listeners' identification of the consonants as dentals steadily increased. After this point in time (Gate 5), identification as dentals began to drop and identification as phonotactically legal consonants (i.e., the velars /g/ and
strains affect how well Chinese subjects perceive the word-final English /it/ vs /d/ contrast. Journal of Phonetics, 17, 299-315.


NOTES

1. These predictions were made using the sum of log frequency of the words in Kubera and Francis (1967) as the predictor because this measure includes information about word, and hence cluster, frequency and the number of words in which each cluster occurs. The same predictions (position independent and word initial) hold when based on either of these measures alone. Although the use of Kubera and Francis, a database of written English, to study the perception of spoken language might not be the preferred choice, this database is one of the only sources that provides information on cluster frequency that is sufficiently large to make meaningful predictions.

2. Graphs containing the full labeling functions from the subanalyses in Experiment 2 can be obtained from the author.

3. The subanalysis was also performed on the data of Experiment 1. Although fewer participants were classified as /l/-dominant or /l/-dominant in each context, the same pattern of results was found. In the /l/ context, only 1 listener did not produce a full labeling function, hearing the continuum mostly as /l/. In the /l/ context, 3 listeners were /l/-dominant, and none was /l/-dominant. In the /l/ context, the /l/-bias was stronger, with 13 /l/-dominant listeners, and none as /l/-dominant. This pattern reversed in the /s/ context, with 6 /l/-dominant and 10 /l/-dominant responders. The fact that half of the listeners in the /s/ context heard the continuum primarily as /l/ or /l/ also explains why the slope of the /s/ function is shallower than the others and why labeling at the /l/ endpoint was higher than that in the /l/ context, although not reliably.

4. In Experiments 1 and 2, obscure identity was shown to affect liquid perception. The reverse can also occur, with liquid identity affecting obscure perception. Massaro and Cohen (1983; Experiment 3) demonstrated this using the categorization paradigm. I found a similar result when listeners were asked to spell naturally spoken CCV syllables presented overhead phones. The stop in phonotactically illegal clusters was frequently misheard as another stop with a different place of articulation (84 listeners were tested). Specifically, /l/ was heard as /l/ 67% of the time but as /l/ only 5% of the time. /l/ was reported as /k/ 13% and as /p/ 45% of the time. See Hale, Segui, Frauenfelder, and Meunier (1998) for related data.

(Manuscript received February 26, 1997; revision accepted for publication July 27, 1997.)