

Running head: STRUCTURAL CONSTRAINTS

Structural constraints in the perception of English stop-sonorant
clusters

Elliott Moreton

Johns Hopkins University

Correspondence address:

Elliott Moreton

Department of Cognitive Science

Johns Hopkins University

Baltimore, MD 21218

moreton@cogsci.jhu.edu

Abstract

Native-language phonemes combined in a non-native way can be misperceived so as to conform to native phonotactics; e.g., English listeners are biased to hear syllable-initial [tr] rather than the illegal [tl] (Massaro & Cohen 1983, Pitt 1998). What sort of linguistic knowledge causes phonotactic perceptual bias? Two classes of models were compared: unit models, which attribute bias to the listener's differing experience of each cluster (such as their different frequencies), and structure models, which use abstract phonological generalizations (such as a ban on [coronal][coronal] sequences). Listeners (N=16 in each experiment) judged synthetic 6x6 arrays of stop-sonorant clusters in which both consonants were ambiguous. The effect of the stop judgment on the log odds ratio of the sonorant judgment was assessed separately for each stimulus token to provide a stimulus-independent measure of bias. Experiment 1 compared perceptual bias against the onsets [bw] and [dl], which violate different structural constraints but are both of zero frequency. Experiment 2 compared bias against [dl] in CCV and VCCV contexts, to investigate the interaction of syllabification with segmentism and to rule out a compensation-for-coarticulation account of Experiment 1. Results of both experiments favor the structure models. (Supported by NSF.)

Speech perception is guided by the expectation that the stimulus is an utterance in the perceiver's language. This finding cuts across every level of language organization: phoneme inventory (e.g., Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975), phonotactics (e.g., Brown & Hildum, 1956), the lexicon (e.g., Ganong, 1980), and syntax (e.g., Garnes & Bond, 1975).

In phonological processing, native sounds combined in a non-native way can be misperceived so as to conform to the native phonotactics (Polivanov, 1931; Sapir, 1933). English listeners transcribe non-native onset clusters as native ones (Brown & Hildum, 1956). Japanese listeners hear an illusory vowel in un-Japanese consonant clusters, and have difficulty discriminating the illusory vowels from real ones (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999). English listeners judging a sound intermediate between [r] and [l] tend to report "r" when the stimulus is [tʔi] and "l" when it is [sʔi] (Massaro & Cohen, 1983, Pitt, 1998). French listeners misperceive even acoustically unambiguous tokens of the non-native word-initial [tl] as the native [kl] in several tasks (Hallé, Segui, Frauenfelder, & Meunier, 1998).

How does native-language phonotactics intervene in the perceptual process? Phonotactic influence is strongest when auditory cues are weakest -- when the stimuli are deliberately made ambiguous, or when the experimental conditions keep the listener from concentrating on the segment in question -- suggesting that linguistic knowledge is used when there are competing parses of the input. This study is concerned with the nature of the linguistic knowledge used to decide between them. Proposals fall into two broad classes.

Unit models attribute perceptual preference for, e.g., [tr] over [tl] to the listener's differing experience of the specific phonological units [tr] and [tl]: One is an attested onset and the other is not (Hallé et al., 1998; Pitt, 1998), one is common and the other is rare (Massaro & Cohen, 1983; Pitt & McQueen, 1998), one is supported by many lexical items which contain it and the other is not (McClelland & Elman, 1986). These models build on a large body of work in other areas of speech processing showing effects of unit frequency (e.g., Vitevich & Luce, 1999; Jusczyk, Luce, & Charles-Luce, 1994; Frisch, Large, & Pisoni 2000; Hay, Pierrehumbert, & Beckman in press) and lexicality (Ganong, 1980; Samuel, 1981; Fox, 1984). They vary widely in details of implementation, but not in ways relevant to the present study.

Structural models state preferences more abstractly, not at the level of specific phoneme combinations but at that of phonological generalizations over classes of phonemes (e.g., Chomsky & Halle, 1968; Prince & Smolensky, 1993). For example, Japanese listeners' tendency to hear [ebzo] as [ebuzo] is derived from a general ban on syllables ending in a non-geminate obstruent, rather than from the specific lack of [bz] sequences in the listener's experience (Polivanov, 1931; Dupoux et al, 1999; Brown & Matthews 2001).

To dissociate these two classes of model experimentally, frequency must be manipulated separately from structural constraints. English stop-sonorant onsets permit this. Both [dl] and [bw] are unattested as syllable onsets in English, as shown in Table 1. Nonetheless, [dl] is commonly classified as "impossible", while [bw] is

"marginal" at worst (Hultzén, 1965; Wooley, 1970; Catford, 1988; Hammond, 1999).

Table 1

Frequency of occurrence of stop-sonorant clusters as onsets in English

Cluster	Position			
	Word-initial		Syllable-initial	
	By types	By tokens	By types	By tokens
Labial				
bw	0	0	0	0
bl	389	27948	890	69100
Coronal				
dw	10	983	16	1003
dl	0	0	0	0
Dorsal				
gw	6	172	59	4834
gl	148	12644	292	19001

Note. Values represent occurrences in the 18.5-million-word London-Lund corpus of written and spoken British English, using the principal pronunciation of each entry in the CELEX EPL.CD lemma database (Baayen, Piepenbrock, & Gulikers, 1995). Phrasal entries (e.g., black-and-blue) were counted as single words. Proper nouns such as Gloucester are included.

There are coherent structural grounds for this difference. Both clusters violate a cross-linguistically widespread constraint against successive consonants with the same place of articulation in the same unit – here, the syllable onset (McCarthy, 1988; Padgett, 1991). The [dl] onset has two successive coronals, while the [bw] onset has two labials. There are two reasons to think that in English, less weight is laid on the labial-labial ban. First, English [r] is labial (Delattre & Freeman, 1968), so the legal, frequent onsets [br pr] violate the same constraint as [bw]. Second, the ban on same-place CC sequences is, cross-linguistically, stronger the more similar the two Cs are in sonority (Selkirk, 1988; Padgett, 1991). Since [l] is less sonorous than [w] (Kahn, 1980; Guenter 2000), the [dl] sequence is closer in sonority than [bw] and hence a worse structural violation. Structural theories therefore predict a larger perceptual bias against [dl] than against [bw].¹

¹English does have [sl], [sn], and [st] onsets, which appear to violate the ban on [coronal][coronal] sequences. However, initial [s] is exceptional in another respect: Unlike all other consonants, it can precede a less sonorous segment. In fact, [s] can be added to any legal onset except a fricative or affricate, and all three-consonant onsets are so formed. The [s] neutralizes the [voice] contrast in a following stop, and palatalizes to [ʃ] before [r], but otherwise does not interact with the rest of the syllable. These facts are ordinarily analyzed by positing a reserved structural slot for [s] at the left margin of the syllable, outside of the onset (e.g., Kenstowicz, 1994:258; Borowsky, 1986:175-179). This account is corroborated by the behavior of the coronal fricative [ʈ], which cannot occupy the [s] slot and thus is subject to the [coronal][coronal] ban: [ʈl], [ʈn], and [ʈt] are impossible onsets.

A unit-based theory which only distinguishes attested from unattested configurations predicts equal biases against [dl] and [bw]. One in which the bias depends on the frequency difference between the legal and illegal endpoint (e.g., [dw] vs. [dl]) would seem to predict a larger bias against [bw] than against [dl]. However, frequency effect magnitudes typically depend on the difference in logarithms of the absolute frequencies (Smith & Dixon, 1971; Rubin, 1976; Hay, et al. in press). As frequency goes to zero, the logarithm approaches negative infinity; hence, the difference in log frequency between [bw] and [bl], or [dw] and [dl], is unboundedly large, and both ought to produce the largest possible phonotactic biases. The differences in log frequencies of [gl gw] are, by comparison, negligible, which allows them to be used as a baseline. Frequency differences between legal stop-liquid clusters did not cause phonotactic perceptual bias in the studies of Mann and Repp (1981) and Pitt (1998).

By measuring the perceptual biases against [dl] and [bw], one can compare the effects of structural constraints when frequency is held constant. This requires that the bias measures be comparable across stimulus sets. The procedure used in most previous work would involve presenting an [l]-[w] continuum in the unambiguous contexts [d_], [b_], and [g_], and measuring the location of the [l]-[w] boundary in each context in terms of continuum steps. This is not possible here, because the [d_] and [b_] contexts are expected to shift the boundary in opposite directions. Different-sized shifts could indicate different-sized phonotactic biases, but they could also simply reflect a closer perceptual spacing of the stimuli at one end of the [l]-[w] continuum, or an uninteresting auditory interaction.

The problem arises from the use of stimulus units to measure the dependent and independent variables. The technique adopted here is to measure the effect of one response on another: Listeners judged a CC cluster in which both Cs were ambiguous, and the dependent measure was the effect of their decision about the first C ("g" vs. "d", or "g" vs. "b") on their decision about the second ("l" vs. "w") (Nearey, 1990). By so doing, one can control stimulus factors completely: The dependence between stop and sonorant judgments can be measured separately for each individual stimulus.

Experiment 1 compares the biases against [dl] and [bw] in CCV syllables. Experiment 2 is a control to insure that the results of Experiment 1 are due to a phonotactic bias, rather than to compensation for coarticulation (Mann, 1980).

Experiment 1

The aim was to measure the dependence of "l"/"w" judgments on "g"/"d" and "g"/"b" judgments in English CCV syllables. All listeners were tested on two separate stimulus sets: an array of stimuli ambiguous among [glæ gwæ dlæ dwæ], and one ambiguous among [glæ gwæ blæ bwæ].

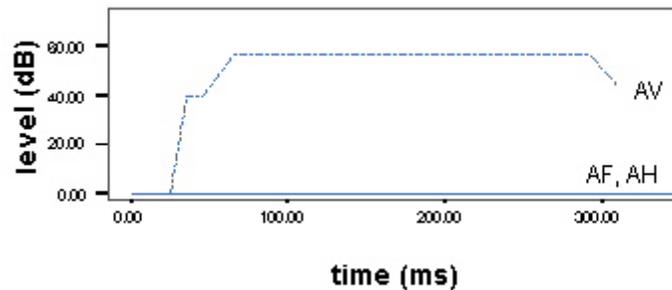
Method

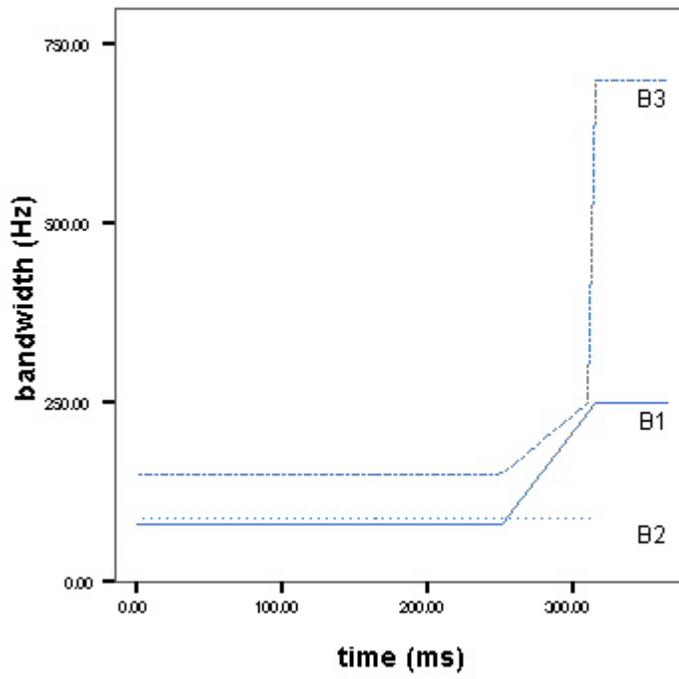
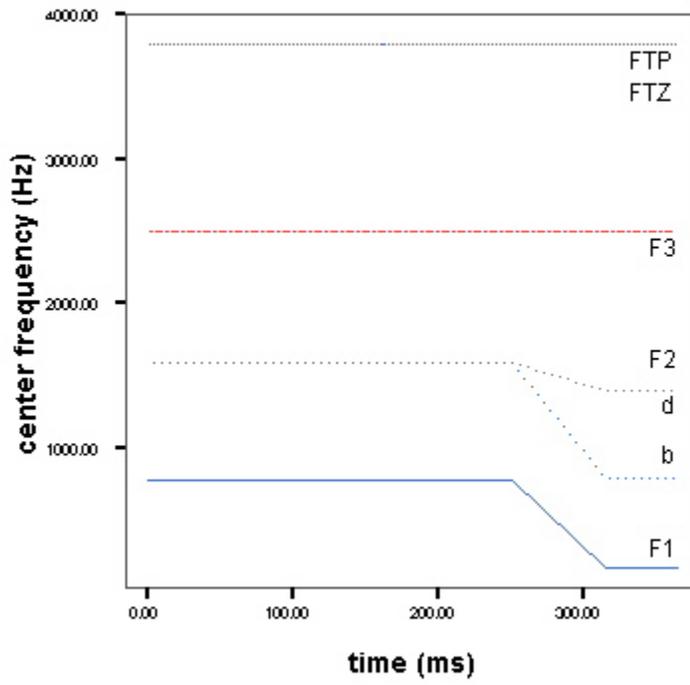
Stimuli

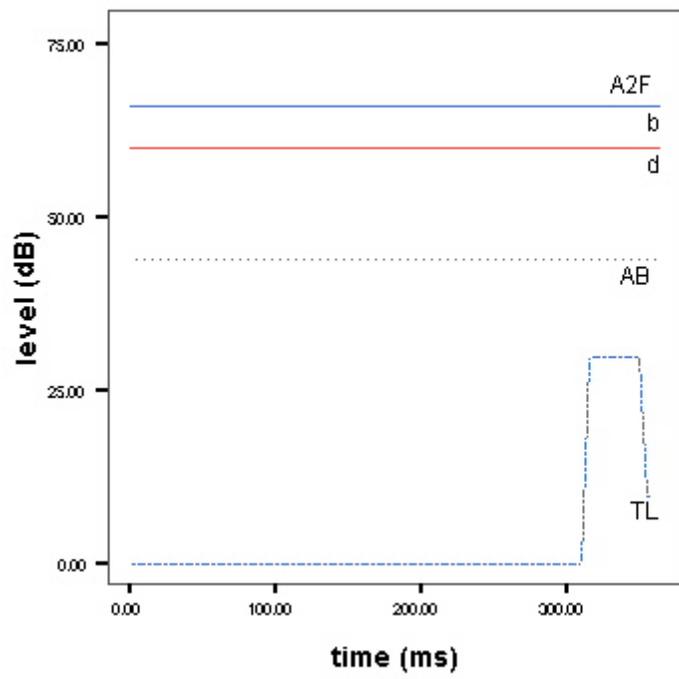
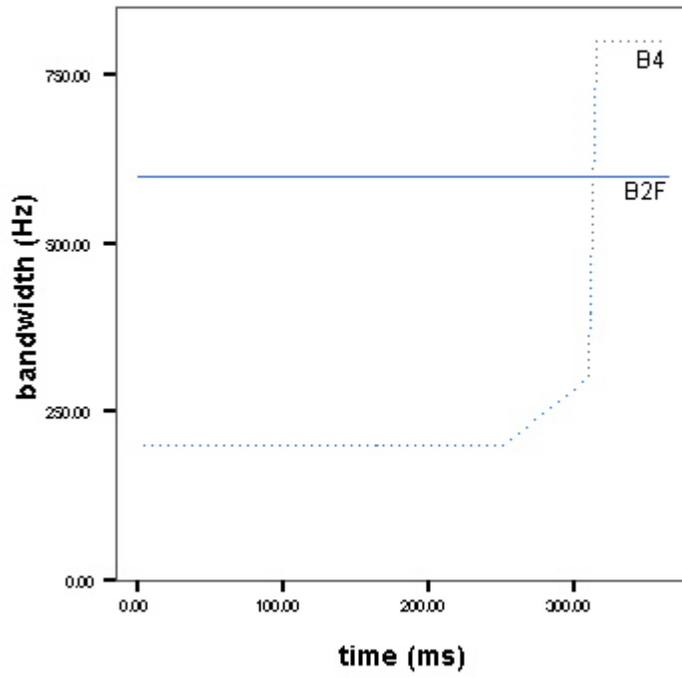
Stimuli were synthetic nonword CCV monosyllables. The V, following Pitt (1998), was [æ]. The second C ranged from [l] to [w]; the first, from [g] to [d] or from [g] to [b]. To prevent listeners from memorizing the individual stimuli, a large stimulus set was used (Crowther & Mann, 1994): six steps along each continuum, making 36 stimuli in each array. Stimuli were identified by a 2-digit code. The first digit specified position on the stop continuum ('0' = most b- or d-like, '5' = most g-like); the second, position on the sonorant continuum ('0' = most l-like, '5' = most w-like).

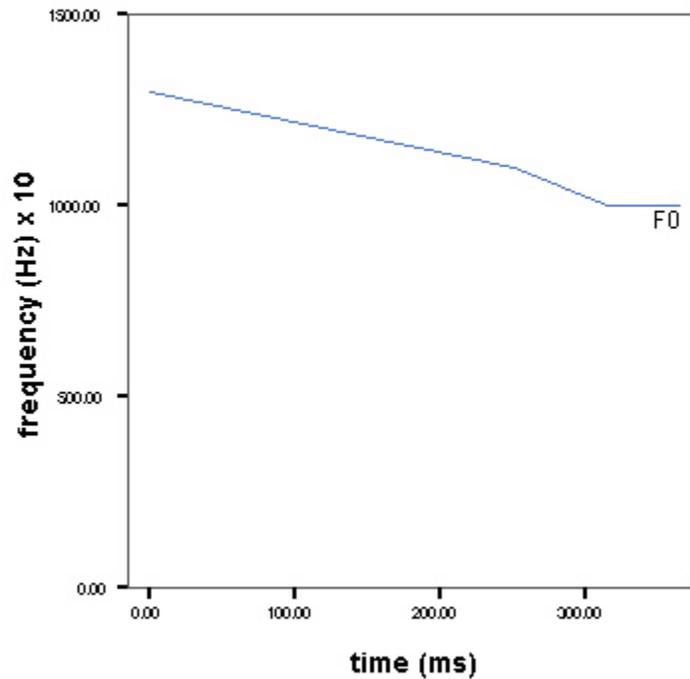
Care was taken to make the stimuli acoustically and articulatorily plausible, and to insure that ambiguous segments were heard as one of the intended phonemes. Synthesizer parameters are shown in Figure 1; only differences between the endpoints are discussed in the text.

Figure 1. Synthesis parameters for the stimuli of Experiment 1. The following parameters, which were the same for all stimuli and did not vary across time, are omitted: GH 50, OQ 30, F6 4900, B6 100, F5 4300, B5 300, F4 3250, F3 2500, FL 20.









The [l] and [w] endpoints followed the acoustic theory of those segments in Stevens (1999, pp. 513–555). The [l] endpoint had a low F2 and high F3, corresponding to an elevated tongue dorsum, and a pole-zero pair near F4, corresponding to the cavity above the tongue blade (Stevens, 1999, p. 545). At the [w] endpoint, the pole-zero pair was absent, and formants above F2 were attenuated to simulate low-pass filtering by a labial constriction. F1 and F2 were lower than in [l], another correlate of labiality.

The stop endpoints differed only in F2 onset, frication bandwidth at F2, and amplitude of the F2 and wide-band frication components. The [g] had a low F2 onset (due to coarticulation with dorsal [l]/[w]) and a compact burst spectrum, with energy concentrated near F2, while [b] and [d] had diffuse burst spectra (Blumstein & Stevens, 1979). The [b]

had the same low F2 onset as [g], while [d] had a higher onset than [b] and less energy in the F2 region. The stop-sonorant transition lasted 65ms.

Parameter values were adjusted to make the endpoint stimuli slightly ambiguous. Intermediate steps were made by interpolating the synthesizer parameters. Interpolation was linear except for the bandwidth of F2 frication, which was interpolated along an exponential curve of the form $B2F = A * \exp(B*r)$, where r went from 0 at the [g] endpoint to 1 at the [b] and [d] endpoints.

Stimuli were synthesized using the cascade branch of the SENSYN terminal analogue synthesizer (Klatt, 1980) with 16-bit resolution, a 16-kHz sampling rate, and a 2-ms frame length. Six formants were used, but only the lowest two varied. Stimuli were low-pass filtered with a sharp cutoff at 5512 Hz.

This procedure yielded two 36-element stimulus arrays: one ambiguous between [glæ gwæ dlæ dwæ] (the "d array"), and one between [glæ gwæ blæ bwæ] (the "b array"). Pretesting with 32 listeners showed that the stimuli sounded natural, were ambiguous, and were only heard as one of the intended syllables.

Participants

Seventeen naive undergraduate native speakers of American English participated as part of a psychology course requirement. None reported speech or hearing deficits. One was dropped for inability to do the practice, leaving 16 valid subjects.

Procedure

Listeners were tested individually in a sound-attenuated booth (IAC Model 401A) during two 15-minute blocks separated by a 5-minute break. Eight listeners heard the "d" array in the first block, while the other 8 heard the "b" array first. In each block, the 36 syllables were presented 5 times in pseudorandom order through Sennheiser EH-1430 headphones. The listener responded by pressing one of four buttons labelled "dw dl gw gl" or "bw bl gw gl". Button order was rearranged between listeners. Listeners had 5 s to respond; the next trial followed after 1 s.

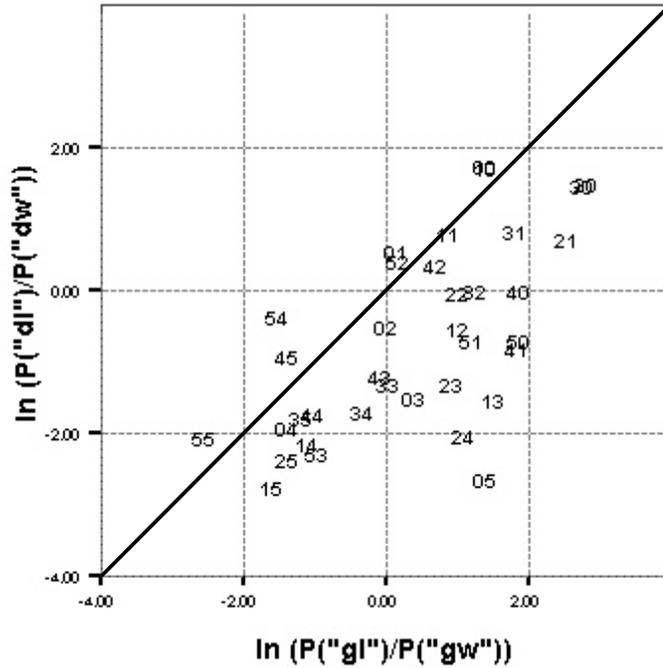
Each block was preceded by a practice without feedback. Each of the four most extreme stimuli (at the corners of the array) was presented three times, for a total of 12 stimuli, in pseudorandom order, and judged by the listener as in the main experiment. The practice was repeated until the listener had used all four responses (accurately or not).

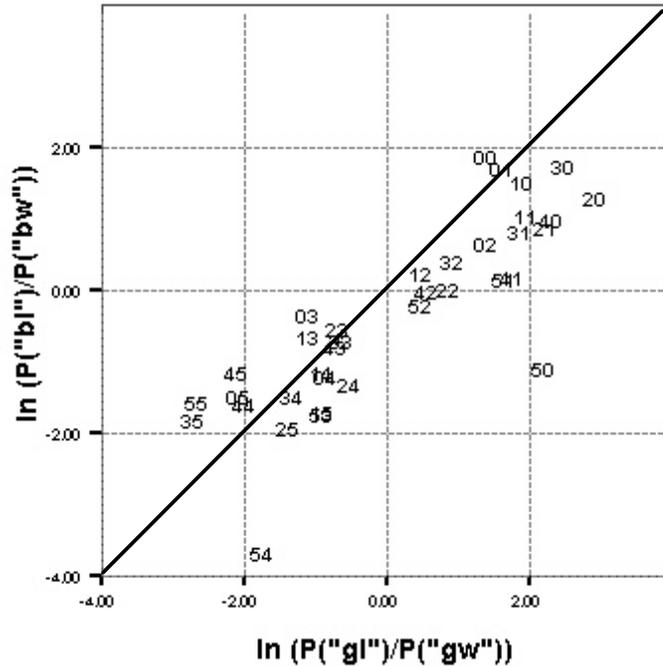
Results

For each stimulus, the 160 responses from all listeners were pooled to estimate the likelihood that it would be put into each of the four categories. The statistic of interest is how the listeners' "l"/"w" judgment on a particular stimulus is affected by their decision about the stop. The "l"/"w" judgment was quantified as the log-transformed odds ratio of "l" versus "w" responses (Macmillan &

Creelman, 1991:15). This was calculated separately for each stop response, as shown in Figure 2. If the stop decision had no effect on the "l"/"w" decision, then all the points in Figure 2 would lie on the line $y = x$. Displacement from this line indicates phonotactic bias.

Figure 2. Log odds ratios for "l"/"w" judgment, contingent on "g"/"d" and "g"/"b" judgment, Experiment 1. Each point represents 16 listeners' pooled responses to one stimulus. Stimulus codes are explained in the text.





For example, Stimulus 33 from the "d" array was judged as "gl" 36 times, "gw" 35 times, "dl" 13 times, and "dw" 72 times. When the stop was identified as "g", the sonorant was equally likely to be classified as "l" or "w": $\ln(P("gl"|\underline{S})/P("gw"|\underline{S})) = \ln(36/35) = 0.028$, plotted on the \underline{x} -axis in Figure 2. When the stop was identified as "d", the sonorant was more likely to be called "w" than "l": $\ln(P("dl"|\underline{S})/P("dw"|\underline{S})) = \ln(13/72) = -1.712$, plotted on the \underline{y} axis. The measure of phonotactic bias is the difference \underline{d} : the log of the "l"/"w" odds ratio contingent on a "g" decision minus that contingent on a "d" or "b" decision, here $0.028 - (-1.712) = 1.740$. The [l] endpoint stimuli (00, 10, ..., 50) evoked mostly "l" responses regardless of the stop judgment and so fell in the upper right-hand corner, while judgments of the [w] endpoints (05, 15, ..., 55) correspondingly fell in the lower left.

For each array, \hat{D} = mean \hat{d} over all stimuli was computed. In the "d" array, \hat{D} = 1.224, indicating that a "d" judgment reduced the odds of "l" by a factor of $\exp(1.224) = 3.40$. In the "b" array, \hat{D} was 0.4762 – an unexpected result, since it means that a "b" judgment, far from reducing the odds of a "w", actually increased them by a factor of $\exp(0.4762) = 1.61$ (corresponding to a clustering of the points below the line $y = x$ in Figure 2b).²

Because the dependent measure, difference in log-odds ratios, bears a complex relation to the individual subject data and is drawn from an unknown distribution, the appropriate statistical test is the non-parametric bootstrap (Efron & Gong, 1983; Efron & Tibshirani, 1993). The null hypothesis $H_0: D = 0$ was tested against the two-sided alternative $H_1: D \neq 0$ using the sensitive procedure recommended by Hall and Wilson (1991). For each array, $B = 10,000$ bootstrap resamples were drawn and used to find \hat{d}_{α} such that $\Pr^* (|\hat{D}^* - \hat{D}| > \hat{d}_{\alpha}) = \alpha$. For the "d" array, $\hat{d}_{0.05} = 0.3986$ and $\hat{d}_{0.01} = 0.5238$. Both are much less than the observed $\hat{D} = 1.224$, allowing rejection of H_0 at the 99% confidence level. For the "b" array, $\hat{d}_{0.05} = 0.4856$ and $\hat{d}_{0.01} = 0.6103$; hence, the \hat{D} of 0.4762 barely misses significance at the 0.05 level.

Discussion

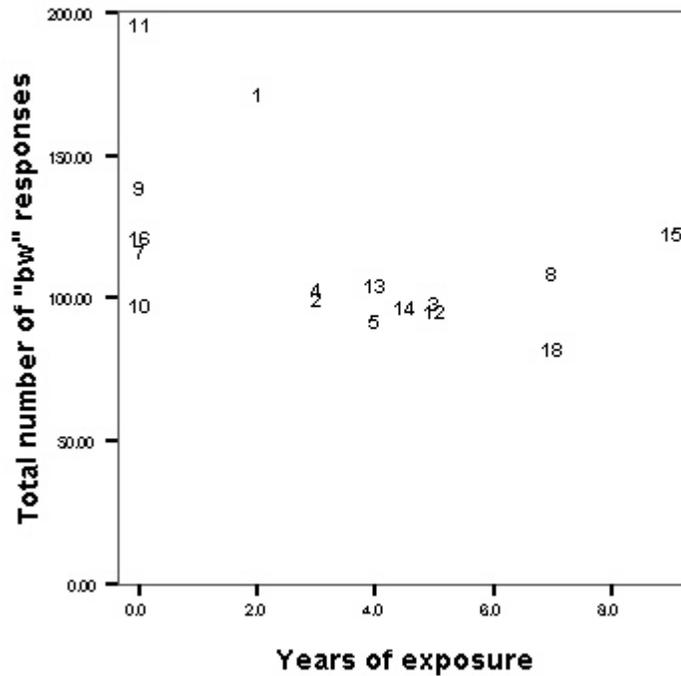
Although both [dlæ] and [bwæ] are unattested in English, a significant phonotactic bias was found only against [dlæ]. This is

²Bias differences are the same for the stop decision conditional on the sonorant decision: $\log (P(gl)/P(gw)) - \log (P(dl)/P(dw)) = \log (P(gl)/P(dl)) - \log (P(gw)/P(dw))$.

consistent with the predictions of the structure models, but not with those of the unit models.

An alternative explanation must be considered. Most of the participants had had up to nine years of exposure to a language with [bw] or [pw] onsets. Could this have allowed them to build perceptual units for these un-English clusters? Each listener's total number of "bw" responses was regressed against years of exposure to French, Spanish, or Mandarin Chinese (see scatterplot in Figure 3). Longer exposure led to slightly fewer "bw" responses. The trend was weak ($R^2 = 0.201$) and due mostly to Listener 11, who had no exposure and a very high rate of "bw" response. When this listener was excluded, the trend vanished ($R^2 = 0.117$). Foreign-language experience does not, therefore, explain the weakness of the bias against [bw].

Figure 3. Total number of "bw" responses in Experiment 1 as a function of individual listeners' exposure to languages containing [bw] or [pw] onsets (French, Mandarin Chinese, or Spanish).



Experiment 2

Another possible source of the effects in Experiment 1 is compensation for coarticulation (Mann, 1980, Mann & Repp, 1981, Mann, 1986). If an ambiguous stop between [d] and [g] is perceived by the listener as [d], it has an atypically low F2 for a [d]. Some of this lowness may be interpreted as labialization spreading from the following consonant. The listener will then be more likely to classify that consonant as the labial [w]. A "d" decision on the stop thus decreases the likelihood of an "l" decision to the sonorant. Because [b] and [g] have similar F2s in this context, the "b"/"g" decision would have a smaller compensation effect. This would produce precisely the observed pattern.

The compensation account can be tested by manipulating the stimuli to alter their phonotactics while leaving their coarticulatory properties intact. As pointed out by Pitt (1998), a cluster which is illegal in an onset may become legal if split by a syllable boundary. A structural account predicts less bias against “dl” responses in [ædlæ] than in [dlæ], because [ædlæ] allows the legal parse [æd.læ]. A compensation account predicts the bias will persist, as compensation has a strong effect across syllable boundaries (Mann, 1980; Mann, 1986; Elman & McClelland, 1988; Pitt & McQueen, 1998), is unaffected by perceived syllabification, and is only slightly reduced, if at all, by a preceding vowel context (Mann & Repp, 1981).

Method

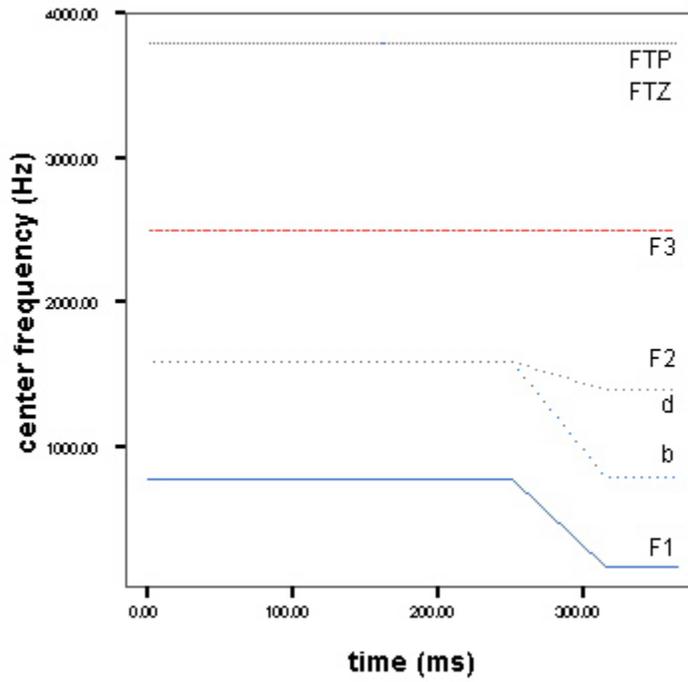
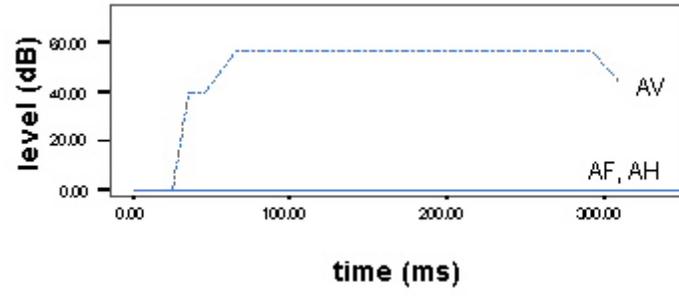
Methods were those of Experiment 1. Only differences will be discussed.

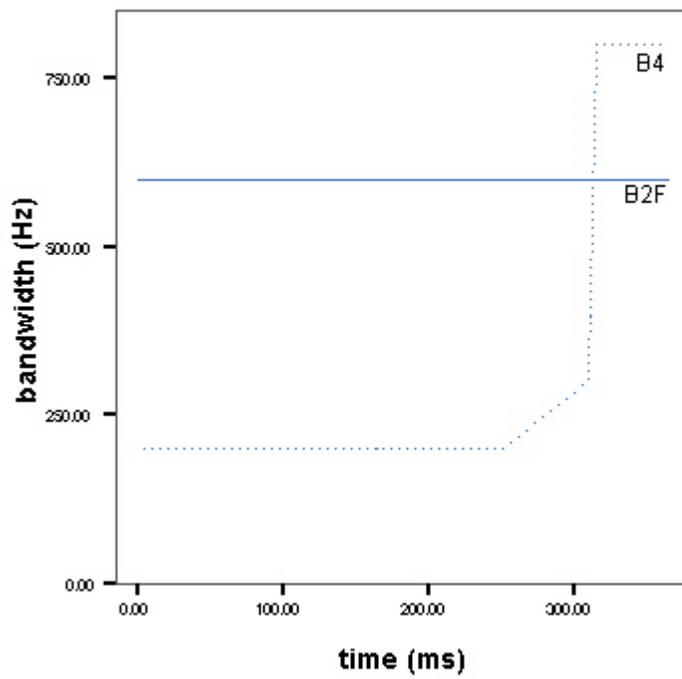
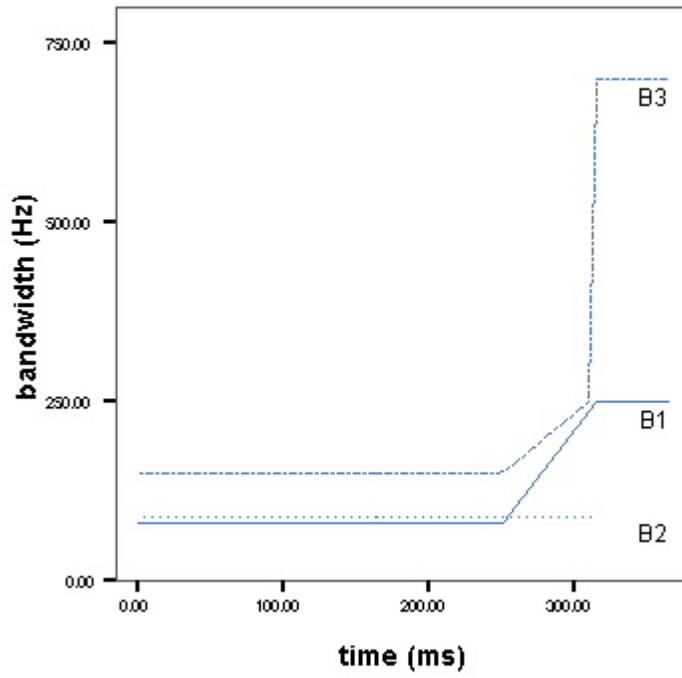
Stimuli

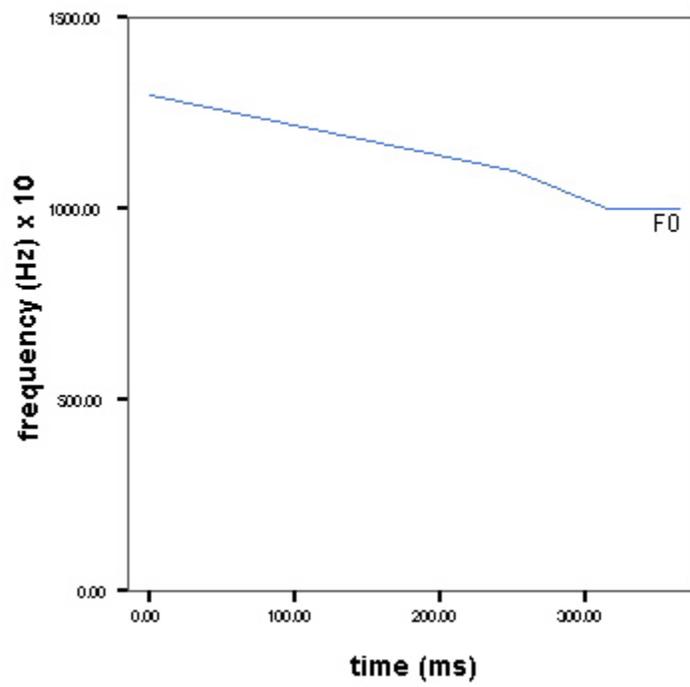
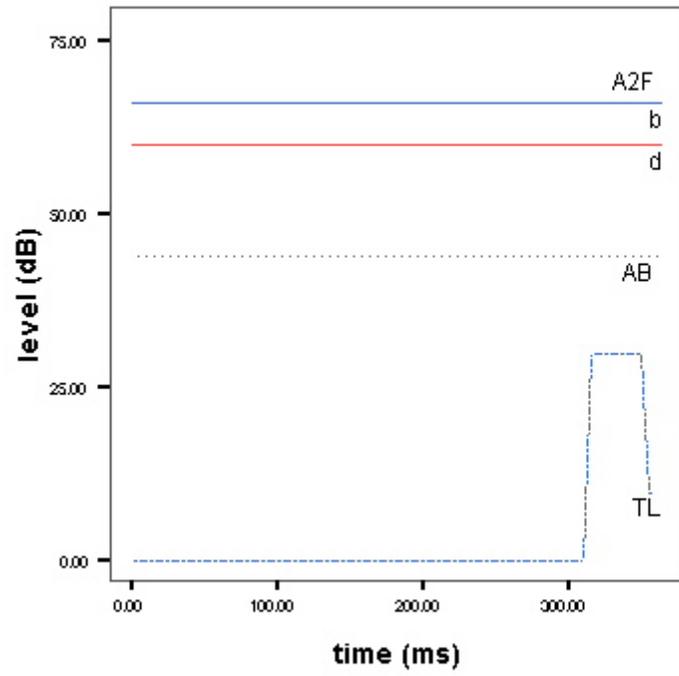
From the endpoints of Experiment 1, a 6-by-6 array of CCV stimuli was constructed, ambiguous between [dwæ dlæ bwæ blæ]. Both [dlæ] and [bwæ] were included to maximize the expected phonotactic effect.

A 6-by-6 array of VCCV stimuli was made by adding a 300 ms [æ] to each of the CCV stimuli. This [æ] used the same parameters as the final [æ], except that F0 began higher (120 Hz). Transition to the stop took 40 ms. A 40-ms voiced closure preceded the release. Details are in Figure 4.

Figure 4. Synthesis parameters for the vowel prefix of Experiment 2.







Participants

Eighteen different members of the same population as in Experiment 1 participated for psychology course credit. Two were dropped because their native language was not English, leaving 16 valid subjects.

Procedure

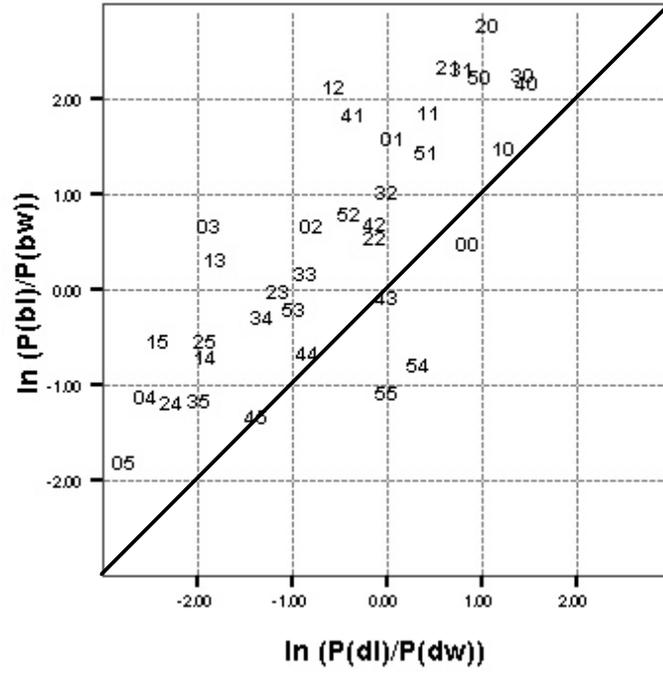
The only difference from Experiment 1 was that all listeners were tested on the VCCV block first and the CCV block second, to avoid priming a V.CCV syllabification.

Results

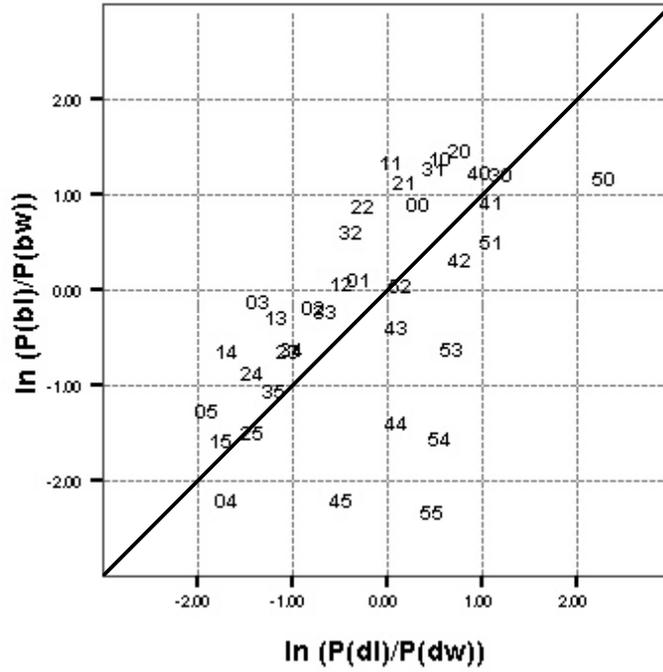
Results are shown in Figure 4. As in Experiment 1, bias appears as displacement from the line $y = x$. The displacement is greater and more consistent in the CCV than the VCCV condition. The test statistic was again D^{\wedge} , the log of the "l"/"w" odds ratio contingent on a "d" decision minus that contingent on a "b" decision, averaged over all stimuli.

Figure 5. Log odds ratios for "l"/"w" judgment, contingent on "b"/"d" judgment, Experiment 2. Each point represents 16 listeners' pooled responses to one stimulus. Stimulus codes are explained in the text.

CCV stimuli



VCCV stimuli



For the CCV array, \underline{D}^{\wedge} is 1.0505, for the VCCV array, it is 0.0648. The same nonparametric bootstrap procedure was used to test significance. For the CCV array, $\underline{d}^{\wedge}_{0.05} = 0.4370$ and $\underline{d}^{\wedge}_{0.01} = 0.5685$, confirming a phonotactic effect. For the VCCV array, the effect did not approach significance: $\underline{d}^{\wedge}_{0.05} = 0.4362$ and $\underline{d}^{\wedge}_{0.01} = 0.6269$.

The results indicate that the bias was eliminated by the availability of a legal parse. This is consistent with a structural account, but not with one based on compensation for coarticulation.

General discussion

Experiment 1 found a perceptual bias against [dlæ], but none against [bwæ]. Unit models predict otherwise, because [dl] and [bw] are both unattested as English syllable onsets. Since listeners' experience of both onsets is identical, that experience cannot explain the difference in performance.

Foreign-language experience also provided no explanation. The difference was not due to auditory factors, since bias was measured separately for each stimulus; rather, it reflected a dependency between the stop and sonorant responses. Experiment 2 confirmed that this dependency was not compensation for coarticulation, because it could be reduced or eliminated by providing a legal parse for the cluster.

It may be objected that listeners' experience of [dl tɪ] and [bw dw] is not in fact identical -- that there is a frequency difference, too small to be detected in an 18-million-word British English corpus, in favor of [bw pw], which university-aged speakers in the United

States are likely to have encountered in foreign place names such as Buenos Aires, southwestern U.S. place names such as Pueblo, or occasional loans like puissant or the colloquial bueno. At a conversational speaking rate of 150 words per minute (Venkatagiri 1999), an 18-million-word corpus would represent only 83 days of continuous speech, or perhaps one to three years of a person's combined input and output. A word occurring less frequently than once in one to three years could escape the corpus -- though an 18-year-old participant in these experiments might have heard it 18 times or more, providing enough experience of [bw pw] onsets to remove the perceptual bias against it. This Undetected Frequency Difference (UFD) Hypothesis is a serious objection, but it is unlikely to be correct.

As has already been pointed out, listeners' acceptance of [bw] was not increased by up to 9 years of explicit training in languages in which [bw pw] onsets are common. It was argued above that this is a ceiling effect; acceptability of [bw pw] cannot be increased by training because the sequences are legal in English. If instead the UFD Hypothesis is correct, then the whole of the gain in acceptability must be caused by the exposure to the first few tokens, with subsequent training having no effect. Hence, it should take exposure to only a small number of tokens to make any sequence legal. But speakers persist in treating some sequences as illegal, even after considerable training (Dupoux et al., 1999; Polivanov, 1931).

In support of the UFD Hypothesis, it may be replied that the listeners were exposed to the undetected low-frequency [bw pw] as children, but received foreign-language training as adults, after the critical period for accentless acquisition. It is certainly true that

infants as young as 9 months are already sensitive to the sound pattern of their language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Friederici & Wessels, 1993; Jusczyk et al., 1994). However, adults can learn phonotactic patterns even without explicit training (Dell, Reed, Adams, and Meyer, 2000). Moreover, a dispreference for [dl tɫ] compared to [bw pw] is found in children who were unlikely to have been exposed to [bw pw] onsets. As shown in Table 2, the midwestern U.S. States of Iowa and Nebraska have a small Hispanic population and almost no place names beginning with [bw pw].

Table 2.

Demographics and occurrences of [bw pw]-initial place names in Iowa and Nebraska (U.S. Census Bureau, 2000; DeLorme, 1998; 2000).

	State	
	Iowa	Nebraska
Population of Hispanic origin	0.0216	0.0462
Place names beginning with		
[bw] (Bue_, Boi_)	Buena Vista (town, county, and college)	--
[pw] (Pue_, Poi_)	--	--

Note: The proportion of the overall U.S. population of Hispanic origin is 0.115. "Buena Vista" is locally pronounced with initial [bw] (Buena Vista College Library staff, personal communication, 2001).

In a study of 1,049 children in Iowa and Nebraska between 2 and 9 years of age, Smit (1993) systematically elicited productions of most of the English word-initial clusters, including [bl pl] and [tw]. The [tw] cluster was sometimes produced as [bw] or [pw], but the [bl] and [pl] clusters almost never became [dl] or [tl], as shown in Table 3.³ This indicates that [d t] are more disfavored before [l] than [b p] are before [w].

³ Similarly, these children also sometimes produce [bl pl] as [bw pw], with no corresponding tendency to turn [tw] into [tl]. Aversion to [tl] may be a contributing factor, but we cannot be sure, because they tend to replace [l] with [w] in all environments.

Table 3.

Errors in production of the initial stop in [bl pl tw] onsets by English-learning children in Iowa and Nebraska (Smit, 1993).

Onset Cluster	Error rate category	
	Occasional	Rare
<u>[tw-] (twins)</u>		
2:0-3:0	f, b	p
3:6-5:6	p, k, d	f, int
6:0-9:0	--	--
<u>[pl-] (plate)</u>		
2:0-3:0	--	--
3:6-5:0	--	b, t
5:6-7:0	--	--
8:0-9:0	--	--
<u>[bl-] (block)</u>		
2:0-3:0	--	--
3:6-5:0	--	--
5:6-7:0	--	--
8:0-9:0	--	--

Note: "Occasional" means "[u]sed by a few groups in an age range with a frequency of 4-10%, or by most groups in that age range at frequencies of 1-4%"; "rare" means "[o]ccurs with a frequency of less than 3%, and only in a few groups in an age range" (Smit, 1993, p.

947). This table includes all errors made by the 1,049 children in the study. "int" = interdental.

The asymmetry is present at the earliest ages tested -- before one would expect most Iowan or Nebraskan children to have had much exposure to Spanish place names. The UFD Hypothesis can therefore only be defended if the perceptual effects of frequency are due chiefly to a very few tokens experienced very early in life. If so, it is an interesting new finding, with many consequences. It implies that, contrary to TRACE, the many words learned after early childhood contribute little to the phonotactic frequency effect. It predicts large individual variation in phonotactics (since the individual is generalizing from a small sample of the adult language, which will necessarily differ more between individuals than a large sample). Finally, it suggests that even large corpora of adult language are inadequate predictors of phonotactic performance, and that research on probabilistic phonotactics should focus more on child-directed speech.

I would argue instead that the present findings are more consistent with a model in which the decision between competing parses is guided by the structural constraints of the perceiver's language - here, the ban on [coronal][coronal] onsets. In Experiment 1, where syllabification was fixed by clear acoustic cues, the choice was between competing CCV parses. The "dl" responses were reduced because a "dl" response could only be supported by the structurally disfavored [dlæ] parse. In Experiment 2, where both segmental identity and syllabification were ambiguous, "dl" responses could be supported by the legal [æd.læ] parse, and the response bias disappeared.

The findings of Pitt (1998, Experiment 2) may be reinterpreted in the same way: "l" response to an [l]-[r] continuum was reduced, relative to a baseline, in the context [mæt_æ], but not in [mæd_æ]. Strong aspiration on the [t] provided an unambiguous cue to V.CCV syllabification (Kirk 2001), allowing only the parses [mæ.træ] and the illegal [mæ.tlæ]. The [mæd_æ] context allowed VC.CV syllabification and thus the legal "l" parse [mæd.læ]. This suggests that prosodic and segmental parse decisions are made in parallel, with the candidate parses representing both phonemes and syllabification: [mæd.læ], [mæ.dlæ], [mæd.ræ], and [mæ.dræ]. The chosen prelexical parse thus provides the essential information for word segmentation and lexical access. Phonotactically impossible parses, such as those with vowelless syllables or illegal onset clusters, are inhibited, leading to the Possible Word Constraint effects observed by Norris, McQueen, and Cutler (1997).

As for the unit models, they fail to predict the lack of bias against [bwæ] for representational reasons: They are unable to connect the legal [br pr] onsets with the unobserved [bw], because they represent linguistic experience in terms of phonemes. Only from a featural viewpoint do [br pr bw pw] form a natural class: [labial][labial] sequences. Listeners' experience of [br pr] evidently affects their processing of [bw], implying that they are using a featural representation. Existing unit models, which are based on phonemic representations, are disconfirmed by the present results.

Author notes

This research was supported by NSF Grant 99-72807 to Paul Smolensky. Many thanks to John Kingston and Lyn Frazier for their advice throughout this study. Mary Beckman, Luigi Burzio, Peter Jusczyk, Terry Nearey, Joe Pater, and the Cognition reviewers also provided invaluable suggestions.

Correspondence concerning this article should be addressed to Elliott Moreton, who is now at the Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218. Electronic mail may be sent via the Internet to moreton@cogsci.jhu.edu.

References

- Baayen, R., R. Piepenbrock, and L. Gulikers (1995). The CELEX lexical database (CD-ROM). Philadelphia: Linguistic Data Consortium.
- Blumstein, S. E., and K. N. Stevens (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. Journal of the Acoustical Society of America 66(4):1001-1017.
- Borowsky, T. J. (1986). Topics in the Lexical Phonology of English. Ph.D. dissertation, University of Massachusetts, Amherst. Amherst, MA: GLSA.
- Brown, Roger W., & Donald C. Hildum (1956). Expectancy and the perception of syllables. Language 32:411-419.
- Brown, C., and J. Matthews (2001). When intake exceeds input: Language-specific perceptual illusions induced by L1 prosodic constraints. Proceedings of the Third International Symposium on Bilingualism, Bristol, U.K., April 18-20, 2001.
- Catford, J. C. (1988). A Practical Introduction to Phonetics. Oxford, U.K.: Oxford University Press.
- Chomsky, N., and M. Halle (1968). The Sound Pattern of English. Cambridge, MA: MIT Press.
- Crowther, C. S., and V. Mann (1994). Use of vocalic cues to consonant voicing and native language background: The influence of experimental design. Perception and Psychophysics 55(5):513-525.
- Delattre, Pierre, and Donald C. Freeman (1968). A dialect study of American R's by X-ray motion picture. Linguistics 44:29-68.
- Dell, G. S., K. D. Reed, D. R. Adams, and A. S. Meyer (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. Journal of

Experimental Psychology: Learning, Memory, and Cognition 26(6):1355-1367.

DeLorme Publishing Company (1998). Iowa Atlas and Gazetteer.
Yarmouth, Maine: DeLorme.

DeLorme Publishing Company (2000). Nebraska Atlas and Gazetteer.
Yarmouth, Maine: DeLorme.

Dupoux, E., K. Kakehi, Y. Hirose, C. Pallier, and J. Mehler
(1999). Epenthetic vowels in Japanese: A perceptual illusion?
Journal of Experimental Psychology: Human Perception and Performance
25(6):1568-1578.

Efron, B., and G. Gong (1983). A leisurely look at the
bootstrap, the jackknife, and cross-validation. The American
Statistician 37(1):36-48.

Efron, B., and R. Tibshirani (1993). An Introduction to the
Bootstrap. New York: Chapman and Hall.

Elman, J. L., and J. L. McClelland (1988). Cognitive penetration
of the mechanisms of perception: Compensation for coarticulation of
lexically restored phonemes. Journal of Memory and Language 27(2):143-
165.

Fox, R. A. (1984). Effect of lexical status on phonetic
categorization. Journal of Experimental Psychology: Human Perception
and Performance 610:526-540.

Friederici, A. D., and J. M. I Wessels (1993). Phonotactic
knowledge and its use in infant speech perception. Perception &
Psychophysics 54:287-295.

Frisch, S. A., N. R. Large, and D. B. Pisoni (2000). Perception
of wordlikeness: Effects of segment probability and length on the
processing of nonwords. Journal of Memory and Language 42:481-496.

Ganong, W. F. (1980) Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance 6(1):110-125.

Garnes, S., and Z. S. Bond (1975). Slips of the ear: Errors in perception of casual speech. Papers from the 11th Regional Meeting of the Chicago Linguistics Society, pp. 214-225. Chicago, Illinois: Chicago Linguistic Society.

Guenter, J. (2000). What is /l/? Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society, University of California, Berkeley, Feb 18-21, 2001.

Hall, P., and S. R. Wilson (1991). Two guidelines for bootstrap hypothesis testing. Biometrics 47:757-762.

Hallé, P. A., J. Segui, U. Frauenfelder, and C. Meunier (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? Journal of Experimental Psychology: Human Perception and Performance 24(2):592-608.

Hammond, M. (1999). The Phonology of English: A Prosodic Optimality-Theoretic Approach. Oxford, U.K.: Oxford University Press.

Hay, J., J. Pierrehumbert, and M. Beckman (in press). Speech perception, well-formedness, and the statistics of the lexicon. Chapter 12 of J. Local, R. Ogden, and R. Temple (eds.), Papers in Laboratory Phonology VI. Cambridge, U.K.: Cambridge University Press.

Hultzén, L. (1965). Consonant clusters in English. American Speech 40:5-19.

Jusczyk, P. W., A. D. Friederici, J. Wessels, V. Y. Svenkerud, and A. M. Jusczyk (1993). Infants sensitivity to phonotactic patterns in the native language. Journal of Memory and Language 32:402-420.

Jusczyk, P. W., P. A. Luce, and J. Charles-Luce (1994). Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language 33:630-645.

Kahn, D. (1980). Syllable-based Generalizations in English Phonology. New York: Garland.

Kenstowicz, M. (1984). Phonology in Generative Grammar. Cambridge, MA: Blackwell.

Kirk, C. J. (2001). Phonological Constraints on the Segmentation of Continuous Speech. Unpublished Ph.D. dissertation, University of Massachusetts, Amherst.

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67:971-995.

Macmillan, N. A., and C. D. Creelman (1991). Signal Detection Theory: A User's Guide. Cambridge, UK: Cambridge University Press.

Mann, V. (1980). Influence of preceding liquid on stop consonant perception. Perception and Psychophysics 28(5):407-412.

Mann, V. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". Cognition 24(3):169-196.

Mann, V., and B. Repp (1981). Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America 69(2):548-558.

Massaro, D. W., and M. Cohen (1983). Phonological context in speech perception. Perception and Psychophysics 34:338-348.

McCarthy, J. (1988). Feature geometry and dependency: A review. Phonetica 43:84-108.

Nearey, T. (1990). The segment as a unit of speech perception. Journal of Phonetics 18:347-373.

Norris, D., J. M. McQueen, and A. Cutler (1997). The possible-word constraint in the segmentation of continuous speech. Cognitive Psychology 34(3):191-243.

Miyawaki, K., W. Strange, R. Verbrugge, A. M. Liberman, J. J. Jenkins, and O. Fujimura (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception and Psychophysics 18:331-340.

Padgett, J. (1991). Stricture in Feature Geometry. Ph.D. dissertation, University of Massachusetts, Amherst. Amherst, MA: GLSA.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. Perception & Psychophysics 60:941-951.

Pitt, M. A., & J. M. McQueen (1998). Is compensation for coarticulation mediated by the lexicon? Journal of Memory & Language 39:347-370.

Prince, A., and P. Smolensky (1993). Optimality Theory: Constraint Interaction in Generative Grammar. MS, Rutgers University.

Rubin, D. C. (1976). Frequency of occurrence as a psychophysical continuum. Perception and Psychophysics 20(5):327-330.

Samuel, A. S. (1981). Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General 110:474-494.

Sapir, E. (1933). La Réalité psychologique des phonèmes. Journal de Psychologie Normale et Pathologique 30:247-265.

Selkirk, E. O. (1988). Dependency, place, and the notion "tier". MS, Department of Linguistics, University of Massachusetts, Amherst.

Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. Journal of Speech and Hearing Research 36:931-947.

Smith, R. C., and T. P. Dixon (1971). Frequency and the judged familiarity of meaningful words. Journal of Experimental Psychology 88(2):279-281.

U. S. Census Bureau (2000). Statistical Abstract of the United States. Washington, D.C.: U.S. Department of Commerce.

Venkatagiri, H. S. (1999). Clinical measurement of rate of reading and discourse in young adults. Journal of Fluency Disorders 24(3):209-226.

Vitevich, M. S., and P. A. Luce (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. Journal of Memory and Language 40:374-408.

Wooley, D. E. (1970). Feature redundancy in consonant clusters. Linguistics 64:70-93.