Bayesian Analysis of Non-native Cluster Production*

Colin Wilson & Lisa Davidson

Department of Cognitive Science, JHU & Department of Linguistics, NYU

1. Introduction

Under a variety of natural and experimental conditions, speakers and listeners distinguish among different types of non-native phonetic/phonological structure. For example, when asked to judge the acceptability of nonce words, speakers exhibit gradient preferences for certain non-native sound sequences over others (e.g., Greenberg and Jenkins 1964, Scholes 1966, Coleman and Pierrehumbert 1997, Vitevitch et al. 1997, Frisch et al. 2000, Treiman et al. 2000, Bailey and Hahn 2001, Frisch and Zawaydeh 2001, Hammond 2004, Shademan 2007, Albright 2009). Similarly, while it has long been known that structures absent from the native language of the listener are particularly susceptible to misperception (or 'perceptual assimilation'; e.g., Massaro and Cohen 1983, Hallé 1998, Pitt 1998, Dupoux et al. 1999), recent research has established that not all non-native structures are misperceived or perceptually assimilated at the same rate (e.g., Moreton 2002, Berent et al. 2007, Haunz 2007, Kabak and Idsardi 2007, Yarmolinskaya 2010).

The goal of the present study is to develop a formal, quantitative framework within which to explain differences in performance on non-native phonetic/phonological structures. Though this framework can be applied to data from acceptability and purely perceptual studies, we focus on analyzing the results from an experiment that involves (at least) both perception and production. In the experiment, native English speakers listened to, and then attempted to produce, nonce words beginning with initial consonant clusters that are legal in Russian but not in English. The elicited productions deviate from the Russian forms in a number of ways; importantly, different non-native clusters — and in some cases even different forms beginning with the same cluster — elicited distinct patterns of performance both in terms of overall error rate and the distribution error types.

* We would like to thank Adam Albright, Edward Flemming, Maria Gouskova, Veronica Monaghan, Kuniko Nielsen, Julia Yarmolinskaya, Paul Smolensky, the participants of NELS 40, and especially Jason Shaw for contributions to and feedback on the research reported here. All errors are ours.

We take the English speakers' productions to reflect phonetic and/or phonological representations formed on the basis of exposure to the Russian materials. According to our analysis, these representations are determined by at least two different types of knowledge. First, participants interpret the fine-grained acoustic/auditory details of the stimulus items using statistical knowledge of how English phonological structures are typically realized. For example, English speakers are more likely to interpret the word-initial cluster of a nonce word as containing a transitional schwa (e.g., *bdagu* → *bᵊdagu*) if the release of the initial consonant exhibits voicing, duration, and other characteristics that place it closer to the range of variation exhibited by English reduced vowels (e.g., Davidson 2006c). Second, participants do not merely attend to low-level properties of the stimulus, but also take into account the gradient grammatical acceptability of possible phonological representations. For example, they are more likely to faithfully represent nonce words beginning with illegal fricative-nasal clusters (e.g., *zmagu*) than nonce words beginning with illegal fricative-stop clusters (e.g., *zbatu*), because the former have higher acceptability according to the English phonotactic grammar.

The two types of knowledge just identified can be quantitatively specified and integrated within a Bayesian approach to perception (e.g., Kersten and Yuille 2003, Norris and McQueen 2008, Feldman et al. 2009). Given a particular stimulus recording {z} produced by a Russian native speaker, the probability that an English listener will represent {z} as [x] is, according to Bayes Theorem, proportional to the probability that [x] would be realized as {z} multiplied by the probability of [x] in English independent of the stimulus:

$$p(\text{representation}=[x] \mid \text{stimulus}=\{z\}) \propto p(\{z\}|[x]) \cdot p([x]),$$

The term p({z}|[x]) indicates where stimulus {z} falls within the native English distribution of auditory realizations of representation [x]. This quantity is determined jointly by the auditorily available characteristics of {z} and the native system of phonetic implementation (e.g., the mental component that determines the probability distribution, in multidimensional auditory space, over realizations of reduced vowels). Considered as a function of the phonetic/phonological representation [x] for a fixed stimulus {z}, this is the *perceptual likelihood* of [x] given {z}. The second term, p([x]), represents the *prior* probability of representation [x]. Though the prior could in a more general analysis be determined by both native phonotactics and patterns of alternation, here we identify p([x]) with the degree of phonotactic acceptability of [x]. The main claim of our proposal is that detailed patterns of performance in experiments of the type analyzed here can only by explained by considering these two terms together, and more specifically by combining them multiplicatively as specified by Bayes Theorem.

The rest of this paper is organized as follows. In section 2 we briefly describe the materials and procedure of the elicitation experiment (for additional details, see Davidson, to appear). Section 3 describes our quantification of perceptual likelihood, identifying a number of measurable properties of the Russian stimuli that we have found to be relevant for predicting the native English speaker's productions (and, indirectly, their perceptions). Section 4 integrates perceptual likelihood with several alternative

models of phonotactic acceptability. Each combination is then assessed with respect to how well it accounts for the quantitative patterns of experimental performance. The results show that the combination of perceptual and phonotactic knowledge is superior to either alone. Among existing gradient phonotactic models, the data supports those based on featural representations and the principle of maximum entropy (e.g., Boersma and Pater 2007, Hayes and Wilson 2008) over alternatives that use only segmental representations (e.g., Vitevitch and Luce 2004) or that weight phonotactic constraints based on frequency of occurrence (e.g., Albright 2009). Section 5 concludes the paper with a brief comparison with alternative theories, and directions for future research.

2. Non-native cluster production data

The data analyzed in this paper come from the production experiment of Davidson (to appear). Participants were 23 native English speakers with no knowledge of Russian or other languages in which the critical word-initial clusters are legal. The materials for the experiment consisted of 240 target nonwords of the form [C1C2aCV] and [C1əC2aCV]; stress was always placed on the first full vowel ([a]) of the word. The 120 [C1C2aCV] stimulus items contained the 60 word-initial consonant clusters specified in Table 1, each cluster occurring with two distinct [aCV] endings. The same consonant sequences and endings were used to construct the matched [C1əC2aCV] items, which differed only in the presence of the schwa. The stimuli were recorded for presentation to the English participants by a native speaker of Russian.

Table 1. Word-initial clusters

Cluster types	Cluster instances	English status
FS	[fp fk ft vb vd vg sp st sk zb zd zg]	all illegal except [sp st sk]
FF	[fs vz sf zv]	all illegal except ?[sf]
FN	[fm fn vm vn sm sn zm zn]	all illegal except [sm sn]
SS	[pt pk bd bg tp tk db dg kp kt gb gd]	all illegal
SF	[pf ps bv bz tf ts dv dz kf ks gv gz]	all illegal
SN	[pm pn bm bn tm tn dm dn km kn gm gn]	all illegal

Key: F = fricative, S = oral stop, N = nasal stop

In each trial of the experiment, a participant heard one of the stimulus recordings repeated twice consecutively and then produced the item aloud. The entire experiment consisted of two blocks, one in which only auditory stimuli were presented and another in which each auditory stimulus was paired with a consistent orthographic stimulus; the order of the two blocks was counterbalanced across participants. Our analysis focuses on the production responses in the auditory-only blocks, as these provide the clearest evidence of specifically phonetic/phonological knowledge and processing. The elicited productions were examined in Praat and coded as either correct (i.e., as having the same phonetic characteristics as the Russian model) or as containing one or more errors. The procedure by which errors were identified and coded was consistent with earlier work (e.g., Davidson 2006ab). The coding system distinguishes *prothesis* (insertion before the

initial cluster, as in *zmagu* → *ᵊzmagu*) from *epenthesis* (insertion within the cluster, as in *bdagu* → *bᵊdagu*), and includes a number of additional error types: C1 deletion, production of C1 as syllabic, featural changes C1 or C2, and metathesis, among others.[1]

Essentially all of the [C1əC2aCV] items were produced without error by the experimental participants. This is unsurprising, since the consonants and other properties of these stimuli make them highly similar to legal English words. The following discussion and analysis therefore focuses on productions of the [C1C2aCV] items, most of which begin with clusters that are not legal in English. Table 2 summarizes the productions of these items in the audio-only condition of the experiment. Each column contains the total count of a particular response type, with the corresponding proportion given in parentheses. Note that the values in the table result from aggregating across all participants and all codable productions (for a total of $N = 1233$ data points).

Table 2. Counts (proportions) of coded production responses to [C1C2aCV] items

correct	prothesis	epenthesis	C1 deletion	other
614 (.50)	82 (.07)	363 (.29)	84 (.07)	90 (.07)

For reasons of space, in this paper we analyze only the patterns of correct productions, prothesis, epenthesis, and C1 deletion (93% of the data). The Bayesian model we develop below can be extended straightforwardly to all response types.

3. Quantification of perceptual likelihood

Previous discussions of non-native consonant cluster perception and production have concentrated primarily on phonetic/phonological properties of the clusters that are relatively general, in the sense that they abstract away from the fine-grained details of particular utterances. These properties include the articulatory gestures that make up the cluster and the coordination relations among those gestures (e.g., Davidson 2006ab, to appear), the sonority contour of the cluster (e.g., Broselow and Finer 1991, Berent et al. 2007, Yarmolinskaya 2010), whether the cluster is legal according to the native grammar (e.g., Hallé et al. 1998, Pitt 1998, Dupoux et al. 1999), whether the cluster can be syllabified by the native grammar (e.g., Kabak and Idsardi 2007), and the frequency with which the cluster occurs in various contexts in the native lexicon (e.g., Davidson et al. 2004, Davidson 2006a, to appear, Berent et al. 2007). Previous studies have also identified a number of grammatical constraints that could affect performance on non-native clusters (e.g., Moreton 2002, Davidson 2006b, Fleischhacker 2005, Zuraw 2007). These constraints often have plausible bases in general properties of speech perception and production (e.g., Ohala and Kawasaki-Fukumori 1997, Hayes et al. 2004). However, like the properties just mentioned, the constraints are stated at a level that abstracts away from many of the details in particular utterances or recorded stimuli.

We agree that general phonetic/phonological properties of clusters, and grammatical constraints that evaluate them, are important for explaining performance on

[1] On the phonetic properties of the insertions, transcribed here as [ᵊ], see Davidson (2005, 2006a).

non-native sequences. Indeed, in the following section we consider several alternative theories of phonotactic acceptability, all of which are stated at this abstract level. However, certain results of the current production experiment lead us to consider another, rather more concrete source of explanation as well. In this section we identify fine-grained acoustic/auditory properties of individual stimulus recordings that appear to be reflected in the participants' responses, and then discuss how their influence can be formalized with the Bayesian concept of perceptual likelihood.

Recall that the word-initial clusters examined in the experiment fall into a small set of types (e.g., fricative-nasal or FN; see Table 1), and that there were two stimulus items beginning with each cluster. A purely abstract phonetic/phonological theory of the data would predict that stimulus items beginning with clusters of the same type — or at the very least items beginning with the same cluster — should elicit similar response patterns. This is appropriate as a first approximation of the actual findings, but detailed examination of the data reveals a number of clear counterexamples. For example, consider the pattern of responses to the four stimulus items of type FN that begin with the voiced fricative [z], as shown in the table below.

Table 3. Proportions of production responses to zN stimulus items

stimulus item	correct	Prothesis	epenthesis	C1 deletion
zmafo	.75	**.13**	.13	.00
zmagu	.36	**.64**	.00	.00
znafe	.78	**.00**	.22	.00
znagi	.45	**.55**	.00	.00

None of these items elicited productions in which C1 was deleted, as would be expected from the salience and associated special status of sibilant fricatives in consonant clusters (e.g., Morelli 1999, Steriade 2001). Apart from this commonality, the response patterns for different items beginning with the FN type, and even different instances of the same cluster, vary widely. We have specifically highlighted the different rates of prothesis. Notice that, for both [zn] and [zm], the proportion of prothesis responses elicited by one stimulus item is more than .50 greater than the proportion elicited by the other; in other words, more than half of the responses to the two items beginning with what is abstractly the same cluster differ in this respect. (The other response proportions also differ, as they must since all the proportions for a stimulus necessarily sum to 1.0).

The within-type and within-cluster differences in Table 3 are not isolated, but part of a larger pattern in which certain stimulus items beginning with voiced fricatives (and to lesser extent voiced stops) elicit many prothesis responses while other phonetically/phonologically matched items elicit very few or none.[2] And such divergences are not limited to prothesis. The rate of C1-deletion is high in some items beginning with SN (stop-nasal) clusters, particularly those with homorganic [bm] and

[2] As another example, the proportion of prothesis responses to the item *vnali* was .40, whereas no prothesis responses were elicited by *vnake*. Similarly, prothesis occurred in 50% of the responses to items *zbatu* and *zgame*, but less often to *zbasi* (33%) and *zgano* (22%).

[dn]; for example, the proportion of C1-deletion responses to *bmalu* is .44 and that to *dnape* is .67. But other items beginning with the same homorganic clusters elicited either no instances of C1 deletion (e.g., *bmada*, .00) or showed much lower rates of this response (e.g., *dnala*, .11). Even epenthesis, which is the most common type of error and the one that is found for all clusters, exhibits rate variations that would be unexpected on purely abstract phonetic/phonological grounds. Striking instances are found among SN items in which the stop and nasal are not homorganic. For example, the epenthesis proportion for *bnapa* was .80, but only .33 for *bnase* (which elicited some instances of C1-deletion but mostly correct productions). 100% of the responses elicited by stimulus item *dmaka* involved epenthesis, but only 60% of responses elicited by *dmafo* showed the same repair (the remaining 40% were coded as correct).

It is unlikely that any of the differences discussed above are due to other abstract properties of the stimulus items, such as the segments, features, or gestures of their CV endings. A more serious concern is that the differences simply reflect noise (unpredictable variation) in the English speakers' perceptions or productions of the non-native clusters, or in the code system. However, we have found that a substantial portion of the within-type and within-cluster variation can be predicted from a small number of acoustic properties of the stimulus recordings. In particular, our analysis incorporates the following properties:

- *Release duration and voicing*. The Russian speaker's productions of most word-initial consonant clusters beginning with stops featured a clear release of the stop (C1), consistent with previous studies using similar materials (e.g., Davidson 2006a). The duration of the C1 release, if present, was measured from the waveform and spectrogram of each stimulus recording; note that this measurement included the burst of C1 and any following aperiodic energy. Additionally, the release was coded as voiced or voiceless; in practice, this was identical to the phonetic/phonological voicing specification of C1 itself. The rationale for these measurements is as follows. Releases that are longer in duration, or voiced, are plausibly more acoustically similar to reduced vowels than shorter or voiceless releases. This leads to the expectation that English speakers will make epenthesis errors more often in response to stimuli beginning with stops that have longer and/or voiced releases. Descriptive non-parametric correlations bear out this expectation: the correlation between C1 release duration and epenthesis rate is highly significant for voiced stops (Kendall's $\tau = .50$, $p < .001$, $N=60$) and also significant for voiceless C1s ($\tau = .33$, $p < .05$, $N=60$).

- *Relativized burst/release amplitude*. The amplitude of the burst and following aperiodic energy of a C1 stop, if present, was quantified in a way similar to that of Stoel-Gammon et al. 1994 and Sundara 2005. The stimulus was first high-pass filtered at 300Hz to remove the effects of the fundamental. The peak amplitude of the burst was then extracted (in dB units) and relativized by subtracting the peak amplitude of the immediately following speech signal (e.g., the closure portion of the following oral or nasal stop). The purpose of this measurement is to distinguish stops with bursts that clearly indicate their presence from stops with bursts that are weak enough to be masked by the following context. Unlike fricatives, whose internal perceptual cues are robust,

stops depend largely upon their burst and release cues to be perceived (e.g., Wright 2004). Therefore, a stop with a weaker burst is predicted to be more susceptible to deletion errors (and, in a more extensive analysis, to feature changes as well). This goes some way toward explaining the finding that deletion of stops in SN clusters, and in particular homorganic SN clusters, occurs at a higher rate than deletion of stops in SS clusters: the C1 bursts in the former are generally lower in amplitude than the following nasal closure, whereas those in the latter are generally higher in amplitude than the following oral stop closure.

- *Prevoicing*. Russian voiced oral stops and fricatives typically exhibit substantial vocal fold vibration in word-initial position, unlike the obstruents traditionally classified as voiced in English. There is one species of word-initial voicing in the Russian stimuli, which we will refer to as *prevoicing*, that is highly predictive of the rate of prothesis errors. Recall from Table 3 that the stimulus item *zmagu* elicited many more prothesis responses than the phonetically/phonologically matched item *zmafo*. Though both of these tokens show phonetic voicing throughout the initial [z], only *zmagu* has prevoicing as we define it: a period of modal voicing that begins early in the sound and that is followed by irregular voicing (which is more characteristic of voiced fricatives) or no vocal fold vibration at all (as often occurs in the middle of word-initial Russian voiced stops). Pitch and sometimes even formant structure can be clearly tracked during the modally voiced portion, unlike the following portion of the same fricative or stop. Because modal voicing, pitch, and formant structure are also characteristic of vowels, stimuli with prevoiced C1s are more likely to be misperceived (and hence misproduced) as containing an initial reduced vowel than stimuli with non-prevoiced C1s. This accounts for two patterns in the production data. First, prothesis errors occur very rarely (< 10 instances) in productions of stimuli that begin with voiceless sounds. Second, the presence vs. absence of prevoicing correctly distinguishes the stimuli in Table 3 that elicit higher rates of prothesis from those that elicit lower rates; this extends to the other cases of within -type and -cluster variation in prothesis rate that were mentioned above. Figure 1 presents an example of prevoicing, which is present within the highlighted portion of the word-initial [z] in *zmagu*.

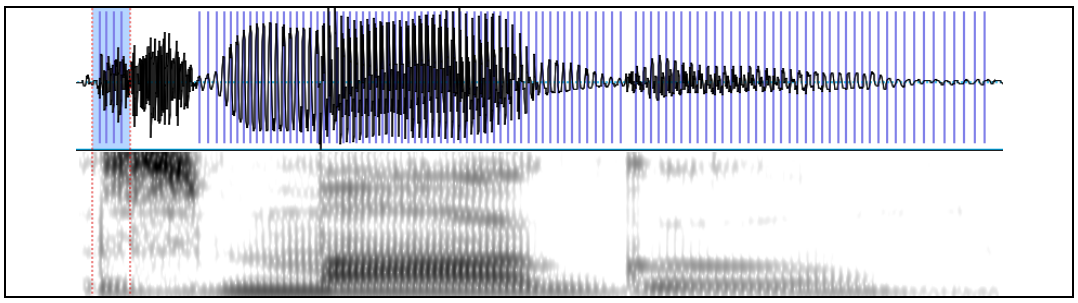

Figure 1. Waveform and spectrogram of stimulus item *zmagu*

Within the Bayesian framework outlined in section 1, auditory encodings of acoustic properties such as these influence performance via the perceptual likelihood function. If recording {z} has characteristics that make it more auditorily similar to typical realizations of phonetic/phonological representation [x] according to the native

system of phonetic implementation, then the perceptual likelihood of [x] given {z} is increased. Here we adopt a provisional formalization of perceptual likelihood according to which each of the properties is related to one candidate representation through an exponential transformation, as in Table 4 below. The β values in the table are free parameters that were fit to the data as part of the modeling in the next section.

Table 4. Relating acoustic properties to perceptual likelihoods

p([ᵊC1C2aCV] | {C1C2aCV})	=	$\exp(\beta_1 \cdot$ C1_prevoicing$)$
p([C1ᵊC2aCV] | {C1C2aCV})	=	$\exp(\beta_2 \cdot$ C1_release_duration$)$ if C1 is voiceless
	=	$\exp(\beta_3 \cdot$ C1_release_duration$)$ if C1 is voiced
p([C2aCV] | {C1C2aCV})	=	$\exp(\beta_4 \cdot$ C1_relative_burst_amplitude$)$

Note: p([C1C2aCV] | {C1C2aCV}) was set to the arbitrary value of 1.0 for all stimuli.

4. Combining perceptual likelihood and phonotactic probability

We began this paper by reviewing the evidence that native speakers make distinctions, in both their acceptability judgments and their perceptual and production errors, among different non-native phonetic/phonological structures. Distinctions of this sort are found throughout the production data analyzed here (see Davidson, to appear, for a complete presentation of the error patterns). As discussed in detail in the previous section, even instances of phonetically/phonological similar or identical clusters were in some cases associated with different error patterns. Differences among the abstractly defined cluster types are also found. For example, the English participants had higher error rates on non-native clusters beginning with voiced obstruents (.64) than on those beginning with voiceless obstruents (.35). Error rates were higher for non-native clusters beginning with stops (SS=.58, SF=.55, SN=.55) than for non-native clusters beginning with fricatives (FS=.45, FF=.40, FN=.37). Furthermore, different non-native cluster types elicited different distributions of errors. For example, epenthesis errors occurred at a higher rate for SS clusters (.56) and SN clusters (.45) than for SF clusters (.28) or any of the non-native clusters beginning with fricatives (FS=.27, FF=.18, FN=.22); in contrast, prothesis errors occurred more often in the productions of fricative-initial clusters.

In this section, we compare a number of analyses of the experimental data, all of which incorporate the model of perceptual likelihood developed above and which are therefore sensitive to fine-grained acoustic/auditory properties of individual recordings. The analyses differ with respect to the other component of our Bayesian approach, namely knowledge of phonotactic probability. Table 5 below illustrates how predicted response patterns are predicted from the combination of perceptual likelihood and phonotactic knowledge, using the stimulus item *zmagu* as an example. (Note that for reasons of space, we have omitted the ending *gu* of the stimulus and the corresponding final syllable [gu] that is shared by all of the responses considered here.)

Table 5. Calculation of predicted response probabilities for recording {zma(gu)}

[x]	perceptual likelihood	phonotactic probability	p([x] | {zma})
[zma]	p([zma] | {zma})	p([zma])	p([zma] | {zma})·p([zma]) / Z
[ᵊzma]	p([ᵊzma] | {zma})	p([ᵊzma])	p([ᵊzma] | {zma})·p([ᵊzma]) / Z
[zᵊma]	p([zᵊma] | {zma})	p([zᵊma])	p([zᵊma] | {zma})·p([zᵊma]) / Z
[ma]	p([ma] | {zma})	p([ma])	p([ma] | {zma})·p([ma]) / Z

where Z = p([zma]|{zma})·p([zma]) + p([ᵊzma]|{zma})·p([ᵊzma]) + p([zᵊma]|{zma})· p([zᵊma]) + p([ma]|{zma})·p([ma])

We considered several models of phonotactic probability: (i) a null model that makes no phonotactic distinction between native and non-native structures; (ii) a binary model that assigns the same low probability to all non-native structures; (iii) a model in which the probability of a structure is equal to the sum of the position-specific probabilities of the segments that it contains (Vitevitch and Luce 20004); (iv) a revised version of the maximum entropy model of phonotactics and phonotactic learning proposed in Hayes and Wilson 2008[3]; (v) the maximum entropy model proposed in Boersma and Pater 2007 (see also Pater et al. 2008), which differs primarily from Hayes and Wilson's model in having a much larger set of phonotactic constraints learned by a different method; and (vi) the phonotactic model developed in Albright 2009, which learns constraints through minimal generalization and weights them according to frequency of occurrence in the native corpus (here, a list of English onsets with their lexical type frequencies).[4] Model (iii) employs segmental representations only, whereas models (iv)-(vi) can form phonotactic generalizations with features (or, equivalently, natural classes). Among the feature-based models, those based on maximum entropy assign weights to constraints in a way that (asymptotically) maximizes the probability of the native corpus. There is no straightforward relationship between constraint weights and the probability of the corpus or related quantities in the model of Albright 2009.[5]

Table 6 below reports the maximum log probability of the production data that is achieved by combining perceptual likelihood with each of the phonotactic models (ii)-(vi). By way of reference, the maximum achieved with the null phonotactic model, which makes no distinctions among word-initial clusters, is -506.70. The third column of the table gives the results of log-likelihood ratio tests, each referred to a χ^2 distribution with one degree of freedom, comparing the non-null phonotactic models with the null. All of

[3] The revised model employs the mathematically motivated *gain* criterion of Della Pietra et al. 1997 to select constraints, not the O/E criterion provisionally adopted by Hayes and Wilson 2008.

[4] Thanks to Adam Albright for making the implementation of Albright 2009 available to us.

[5] Under all of the models considered in the text, the treatment of phonotactic probability was simplified as follows: all responses of the form [ᵊC1C2aCV] were assigned probability π_{proth}, all responses of the form [C1ᵊC2aCV] were assigned probability π_{ep}, and all responses of the form [C2aCV] were assigned probability π_{del}. The three π parameters were optimized for each model separately. Furthermore, probabilities of [C1C2aCV] responses were calculated with models (i)-(vi) by evaluating only the word-initial consonant cluster; these values were divided by a smoothing parameter *T* as in Hayes and Wilson (2008:399). The total number of free parameters for the perceptual likelihood and phonotactic models was therefore 8 = 4 (β_1, β_2, β_3, β_4) + 3 (π_{proth}, π_{ep}, π_{del}) + 1 (*T*), which compares favorably with the number of data points being predicted (*N* = 480, i.e., 4 response possibilities for each of 120 [C1C2aCV] stimuli).

the models except (iii), due to Vitevitch and Luce 2004, and (vi), due to Albright 2009, are significantly better than the null model at the $\alpha = .01$ or $\alpha = .001$ level. The highest data probabilities are achieved by the two maximum entropy models (which are not distinguished from one another by this data). Further tests, omitted for reasons of space, establish that phonotactics alone cannot explain the patterns of production responses: perceptual likelihood, with its sensitivity to acoustic/auditory characteristics of individual recordings, is a necessary component of the analysis. But perceptual likelihood does not supplant phonotactics: the two types of knowledge each make significant contributions.

Table 6. Assessment of perceptual likelihood and phonotactic combinations

phonotactic model	log data prob.	LRT against null (i)
(ii) binary	-497.25	$\chi^2(1) = 18.89$, $p < .001$
(iii) summed unigram	-514.00	$\chi^2(1) = 2.16$, n.s.
(iv) Hayes and Wilson 2008	-487.41	$\chi^2(1) = 28.73$, $p < .001$
(v) Boersma and Pater 2007	-486.57	$\chi^2(1) = 29.57$, $p < .001$
(vi) Albright 2009	-513.00	$\chi^2(1) = 3.14$, $p < .10$

The figure below is a scatterplot of the relative frequency of each response type, for each of the 120 critical stimuli, against the probability that is predicted by combining perceptual likelihood with phonotactic model (iv). The plotting symbol '+' indicates a correct responses, '×' indicates prothesis, 'Δ' indicates epenthesis, and 'o' indicates deletion. Stimulus labels are included to provide a sense of which sorts of items are poorly fit by the current version of the model. To take one example, the rate of C1 deletion for items beginning with homorganic stop-fricative combinations, such as *pfama* and *tsapi*, is under-predicted. Over all, however, there is a clear positive relationship between the predictions and the data (Kendall's $\tau = .64$, $p < .001$, $N = 480$).

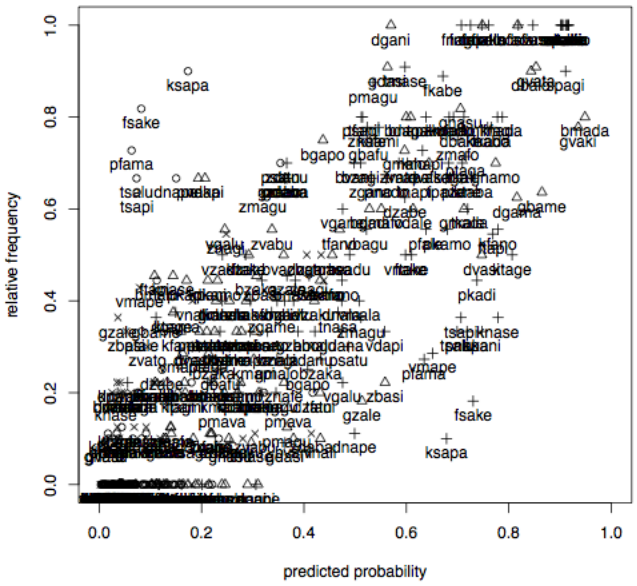

Figure 2. Predicted and observed response rates

5. Conclusion

In this paper, we have shown that detailed patterns in English speakers' productions of non-native word-initial clusters can be accounted for within a Bayesian framework that integrates probabilistic knowledge of phonetic implementation and phonotactic well-formedness. Our analysis has identified specific acoustic parameters that contribute to perceptual likelihood, including burst and release properties as well as prevoicing, and has shown how different phonotactic models can be meaningfully assessed and compared with respect to production data. In accord with previous findings (e.g., Coleman and Pierrehumbert 1997, Frisch et al. 2000), the results favor phonotactic models that make gradient distinctions among non-native phonetic/phonological structures, rather than simply distinguishing legal from illegal structures. Among existing gradient phonotactic models, the results support those that represent sounds with features (e.g., Albright 2009) and those that derive the weights of constraints from the principle of maximum entropy (e.g., Boersma and Pater 2007, Hayes and Wilson 2008). The present study therefore contributes to a growing body of research that identifies fundamental properties of phonotactic knowledge, and provides an explicit model of how that knowledge, in concert with other factors, affects performance in a particular experimental paradigm.

The approach developed here could be applied to a wide range of data, including acceptability judgments, results of purely perceptual experiments, results of production experiments that present orthographic as well as auditory stimuli (e.g., Vendelin and Peperkamp 2006, Davidson, to appear), and data from naturally-occurring behavior such as loanword adaptation (e.g., Kenstowicz and Uffmann 2006, Davidson 2007). In contrast to previous approaches within this broad domain, ours does not assume that the phonological grammar is solely responsible for adaptations of non-native structures (cf. Hyman 1970), and does not rely on a binary distinction between legal and illegal structures (as appears to be assumed in Peperkamp et al. 2008). Our approach is also distinguished by being explicitly Bayesian, in line with work from other cognitive domains in which detailed information about the stimulus is integrated with prior expectations (cf. the conceptually similar proposals of Berent et al. 2009 and Boersma and Hamann 2009). Further development of the approach will involve addressing inadequacies in the provisional theory of perceptual likelihood (e.g., the failure to predict C1 deletion in certain SF clusters), more direct testing of the predictions of that theory with perceptual experimentation, and incorporation of a more complete phonotactic component that evaluates entire phonological representations.

References

Albright, Adam. 2009. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26:9-41.

Bailey, Todd M. and Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44:568-591.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2007. What we know

about what we have never heard: Evidence from perceptual illusions. *Cognition* 104:591-630.

Berent, Iris, Tracy Lennertz, Paul Smolensky, and Vered Vaknin-Nusbaum. 2009. Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology* 26:75-108.

Boersma, Paul, and Joe Pater. 2007. Constructing constraints from language data: The case of Canadian English diphthongs. Talk presented at *NELS* 38.

Boersma, Paul, and Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In *Loanword phonology*, ed. Andrea Calabrese and W. Leo Wetzels, 11-58. Amsterdam: John Benjamins.

Broselow, Ellen, and Daniel Finer. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7:35-59.

Coleman, John, and Janet Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. In *Computational Phonology, Third Meeting of the ACL Special Interest Group in Computational Phonolog*y, 49–56. Somerset, NJ: Association for Computational Linguistics.

Davidson, Lisa. 2005. Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics* 19:619-633.

Davidson, Lisa. 2006a. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34:104-137.

Davidson, Lisa. 2006b. Phonotactics and articulatory coordination interact in phonology: Evidence from non-native production. *Cognitive Science* 30:837-862.

Davidson, Lisa. 2006c. Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica* 63:79-112.

Davidson, Lisa. 2007. The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24:261-286.

Davidson, Lisa. to appear. Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. In *Journal of Phonetics.*

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25:1568-1578.

Feldman, Naomi H., Thomas L. Griffiths, and James L. Morgan. 2009. The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116:752-782.

Fleischhacker, Heidi. 2005. Similarity in phonology: Evidence from reduplication and loan adaptation. Doctoral dissertation, UCLA, Los Angeles, CA.

Frisch, Stefan A., Nathan R. Large, and David 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.

Frisch, Stefan, and Bushra A. Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77:91-106.

Greenberg, Joseph H., and James J. Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20:157-177.

Hallé, Pierre A., Juan Segui, Uli Frauenfelder, and Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24:592-608.

Hammond, Michael. 2004. Gradience, phonotactics, and the lexicon in English phonology. *International Journal of English Studies* 4:1-24.

Haunz, Christine. 2007. Factors in on-line loanword adaptation. Doctoral dissertation, University of Edinburgh.

Hayes, Bruce, Robert Kirchner, and Donca Steriade, eds. 2004. *Phonetically-based phonology*. Cambridge: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379-440.

Hyman, Larry. 1970. The role of borrowing in the justification of phonological grammars. *Studies in African Linguistics* 1:1-48.

Kabak, Baris, and William Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech* 50:23-52.

Kenstowicz, Michael, and Christian Uffmann, eds. 2006. *Loanword phonology: Current issues*. Special issue. *Lingua* 116:921-1194.

Kersten, Daniel, and Alan Yuille. 2003. Bayesian models of object perception. *Current Opinion in Neurobiology* 13:1-9.

Massaro, Dominic W., and Michael M. Cohen. 1983. Phonological context in speech perception. *Perception and Psychophysics* 34:338-348.

Moreton, Elliott. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84:55-71.

Morrelli, Frida. 1999. The Phonotactics and phonology of obstruent clusters in Optimality Theory. Doctoral dissertation, University of Maryland, College Park, MD.

Norris, Dennis, and James McQueen. 2008. Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review* 115:357-395.

Ohala, John J., and Haruko Kawasaki-Fukumori. 1997. Alternatives to the sonority hierachy for explaining segmental sequential constraints. In *Language and Its Ecology: Essays in Memory of Einer Haugen*, ed. Stig Eliasson and Ernst Hakon Jahr, 343-365. Berlin: Mouton de Gruyter.

Pater, Joe, Elliott Moreton, and Michael Becker. 2008. Simplicity biases in structured statistical learning. Poster presented at the Boston University Conference on Language Development.

Peperkamp, Sharon, Inga Vendelin, and Kimihiro Nakamura. 2008. On the perceptual origin of loanword adaptations: Experimental evidence from Japanese. *Phonology* 25:129-164.

Pitt, Mark. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics* 60:941–951.

Scholes, Robert. 1966. *Phonotactic Grammaticalit*y. The Hague: Mouton.

Shademan, Shabnam. 2007. Grammar and analogy in phonotactic well-formedness judgments. Doctoral dissertation, UCLA, Los Angeles, CA.

Steriade, Donca. 2001. The phonology of perceptability effects: The P-map and its consequences for constraint organization. Ms., MIT.

Stoel-Gammon, Carol, Karen William, and Eugene Buder. 1994. Cross-linguistic differences in phonological acquisition: Swedish and American /t/. *Phonetica* 51:146-158.

Sundara, Megha. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America* 118:1026-1037.

Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tincoff, and Margo Bowman. 2000. English speakers' sensitivity to phonotactic patterns. In *Papers in Laboratory Phonology V: Acquisition and the Lexico*n, ed. Michael B. Broe and Janet Pierrehumbert, 269–282. Cambridge: Cambridge University Press.

Vendelin, Inga, and Sharon Peperkamp. 2006. The influence of orthography on loanword adaptation. *Lingua* 116:996-1007.

Vitevitch, Michael S., and Paul A. Luce. 2004. A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers* 36:481-487.

Vitevitch, Michael S., Paul A. Luce, Jan Charles-Luce, and David Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40:47–62.

Wright, Richard. 2004. A review of perceptual cues and robustness. In *Phonetically-based phonology*, ed. Bruce Hayes, Robert Kirchner, and Donca Steriade, 34-57. Cambridge: Cambridge University Press.

Yarmolinskaya, Julia. 2010. Perception and acquisition of second language phonology. Doctoral dissertation, Johns Hopkins University, Baltimore, MD.

Colin Wilson
Department of Cognitive Science
Johns Hopkins University
Baltimore, MD 21218

colin@cogsci.jhu.edu

Lisa Davidson
Department of Linguistics
New York University
New York, NY 10003

lisa.davidson@nyu.edu