Articulatory and perceptual influences on the production of non-native consonant clusters

Lisa Davidson
Johns Hopkins University, Baltimore, MD, USA
E-mail: davidson@cogsci.jhu.edu

ABSTRACT

One hypothesis regarding speakers confronted with sequences phonotactically ill-formed for their language is that they will treat these structures as comparably "illegal". However, production results show that while English speakers given non-native fricative-initial onset clusters do not produce them with perfect accuracy, they are better when the first consonant (C1) of the cluster is /f/, followed by /z/, and lastly by /v/. These findings can be interpreted in terms of a combination of perceptual and articulatory influences on phonotactic structures. The disadvantage of C1=/f,v/ relative to sibilants /s,z/ is perceptually motivated: low-intensity fricatives may be insufficiently detectable at obstruent cluster edges. However, the disadvantage of C1=/z/ relative to /f/ may be articulatory in origin, since oral pressure buildup in obstruent clusters makes voicing difficult to maintain. That speakers discriminate /f/, /z/, and /v/-initial clusters, which are not related on an intrinsic phonetic difficulty scale, suggests that they are distinguished at a phonological level.

1. INTRODUCTION

Previous research in second language (L2) word-initial cluster acquisition reveals that L2 speakers do not produce all non-native clusters with equal accuracy, even if all of the clusters in question are illegal in the speaker's native language. For example, an investigation of the production of English /pr/, /br/, /fr/, initial clusters by Japanese and Korean speakers, where neither language has /r/ as the second segment of a cluster, showed that while the speakers had relatively little trouble with /pr/, they were less accurate on /br/, followed by /fr/ [1]. These results were explained using the phonological notion of markedness: as elements of the clusters became more marked (voiced stops are more marked than voiceless stops but less marked than fricatives), L2 learners were less likely to produce them correctly.

In another study, the production of 10 different Polish word-initial clusters by English speakers indicated that the target clusters could be divided into "difficult", "intermediate", and "easy" categories [2]. For example, /z/+sonorant clusters were least difficult to produce, voiceless obstruent clusters were in the "intermediate" category, and clusters with an initial voiced non-coronal obstruent were produced least accurately. This study also concluded that as the members of the clusters became more marked on the dimensions of place, voice, and manner of articulation, speakers produced them less accurately.

The analyses of the production data in each of these studies treat place, voice and manner as abstract markedness features. However, insights from recent phonological research suggest that production disparities may be explainable in terms of phonetic (perceptual and/or articulatory) characteristics of the segments and the surrounding contexts [3-6]. In this study, articulatory and perceptual factors and how they relate to the relative markedness of word-initial clusters is addressed with an experiment employing fricative-initial clusters. It is hypothesized that the markedness of certain place and voice features is based on specific phonetic characteristics that distinguish them from the optimal fricative /s/.

2. PHONETIC FACTORS AND THE COMPOSITION OF CONSONANT CLUSTERS

In order to examine the role of the initial segment in determining the phonotactic status of word-initial consonant clusters, English speakers were tested on their ability to produce clusters beginning with /s/, /f/, /z/, and /v/. These clusters are found word-initially in Czech, the language used to record the auditory stimuli for this experiment. Since English has /s/-initial clusters, it is assumed that being fricative-initial per se is not disallowed by the English grammar. Rather, it is the place and/or voicing of /f/, /z/, and /v/ that are prohibited.

With respect to the effect of place, a number of phonetic studies of fricatives have shown that the sibilants /s, z/ are significantly more perceptible than non-sibilants /f, v/ for several reasons: (a) because the sound source is filtered by a front-cavity resonance for sibilants /s, z/, the amplitude of the frication noise is much greater than that for labiodental fricatives, which have no cavity in front of the obstruction [7]; (b) it has been claimed that this high-intensity noise and clear timbre contribute to the salience and distinctiveness of /s/, making its acoustic properties almost vowel-like [8]; (c) alveolar fricatives have a long anterior cavity, resulting in a well-defined, distinct spectral shape [7, 9]; (d) /f, v/ have significantly lower noise amplitude relative to /s, z/ [9]; and (e) discriminant analysis and perceptual confusion tests have shown that non-sibilant fricatives like /f/, /v/, /θ/, and /ð/ are significantly more likely to be confused with one another than are sibilants such as /s/, /z/, /ʃ/ and /ʒ/ [9, 10].

Unlike /f/, the failure of /z/ to appear in word-initial clusters does not seem to be motivated by perceptual reasons. Perceptually, /z/ is nearly as salient as /s/ both in terms of its internal features as well as in comparison to

 ISBN 1-876346-48-5

other fricatives. Instead, voiced fricatives are disadvantaged compared to voiceless ones on articulatory grounds. The optimal situation for obstruent voicing occurs when oral pressure is maximally lower than glottal pressure. For the most favorable frication, however, oral pressure should be maximally higher than atmospheric pressure, thus setting up conflicting articulatory requirements for the production of voiced fricatives [11, 12]. Voiced obstruent clusters are further disadvantaged by the fact that they are longer in duration than single voiced obstruents, which may require the conflicting air pressure requirements to be sustained for longer than the speech system can accommodate [13, 14]. Voiced fricative-stop clusters have articulatory shortcomings similar to those of fricative-fricative clusters; just as frication requires high oral pressure, stops block the flow of air, which likewise causes oral pressure to increase, adversely affecting the pressure drop needed for voicing.

3. EXPERIMENT

The experiment conducted in this study tests the hypothesis that English speakers fail to accurately produce non-native clusters with initial fricatives other than /s/ because (i) the weak-intensity fricatives (like /f/ and /v/) contain poor perceptual cues, and (ii) voiced obstruent sequences (like those beginning with /z/ or /v/) are articulatorily difficult to produce. While these segments may be dispreferred cluster-initial consonants compared to /s/, it is not clear whether English speakers would treat them as equally detrimental, or whether either the weak intensity of fricatives like /f/ or the voicing of fricatives like /z/ makes that segment particularly more unfavorable. This question leads to the prediction of three possibilities with respect to production accuracy based on the effect of the first consonant, shown in (1); these are tested in this experiment.

(1) *Accuracy*: More ➞ Less

Possibility 1—Initial weak-intensity and voiced fricatives are equally disfavored in clusters	/s/ ≻ /f/ = /z/ ≻ /v/
Possibility 2—Initial weak-intensity fricatives are less favorable than voiced fricatives in clusters	/s/ ≻ /z/ ≻ /f/ ≻ /v/
Possibility 3—Initial voiced fricatives are less favorable than weak-intensity fricatives in clusters	/s/ ≻ /f/ ≻ /z/ ≻ /v/

If weak-intensity frication and voicing are equally unfavorable, there should be no difference in production accuracy of /fC/ or /zC/ clusters. This corresponds to Possibility 1. If one or the other has a greater effect, however, then Possibility 2 or 3 could occur. Regardless of whether intensity is more unfavorable than voice or vice versa, performance on /vC/ clusters should always be least accurate, since /v/ is both weak-intensity *and* voiced. Likewise, because English does contain /s/-initial clusters, participants should be at or near 100% accuracy on these.

Participants. The participants were 20 Johns Hopkins undergraduates. All of them were native speakers of English and had no exposure to Slavic languages.

Materials. The target words used in the study were pseudo-Czech words with /s/, /f/, /z/, and /v/-initial clusters. These initial segments were combined with stops, fricatives and nasals to create 24 word-initial clusters. All possible combinations are given in (2). Four distinct CCaCV tokens were created for each onset, for a total of 96 target words. The stimuli were recorded by a native Czech speaker using the Kay Elemetrics Computerized Speech Lab (CSL) at a 44.1-kHz sampling rate.

FIRST SEGMENT	CLUSTER
/s/	/sm/, /sn/, /sf/, /sp/, /st/, /sk/
/f/	/fm/, /fn/, /fs/, /fp/, /ft/, /fk/
/z/	/zm/, /zn/, /zv/, /zb/, /zd/, /zg/
/v/	/vm/, /vn/, /vz/, /vb/, /vd/, /vg/

(2) Pseudo-Czech word-initial clusters

Procedure. All 96 words were presented to each speaker in a repetition task. At the start of each trial, the target word written in English orthography appeared on the screen and remained there for the rest of the trial. Twenty milliseconds after the word appeared on the screen, the target stimulus recorded by a native Czech speaker was presented auditorily to the participant through external speakers. The word was repeated again 300ms after the end of the first auditory presentation. Participants were instructed to listen to the two repetitions of the word, and then repeat it one time into the microphone. They then pressed the space bar to move on to the next trial. Each trial lasted 2500ms. Participants were given five practice trials.

Coding. For each experimental condition, the waveform and spectrogram for each target word were analysed using Praat for Windows [15] to determine what, if any, error had been made. There were six possible response categories for each target: Correct (C), Epenthesis (E), Prothesis (P), Deletion (D), Segment Change (S), and Other (O). The correct productions did not contain any period of voicing, aspiration, or formant structure between the two consonants. However, if any of these phonetic characteristics were present, the target was coded for epenthesis. Repairs are exemplified in (3):

RESPONSE	EXAMPLE
Correct	/zvabu/ → [zvabu]
Epenthesis	/zvabu/ → [zəvabu]
Deletion	/zvabu/ → [zabu], [vabu]
Prothesis	/zvabu/ → [əzvabu]
Seg. Chg.	/zvabu/ → [svabu]
Other	/zvabu/ → ∅, [vəvabu], [sfada]

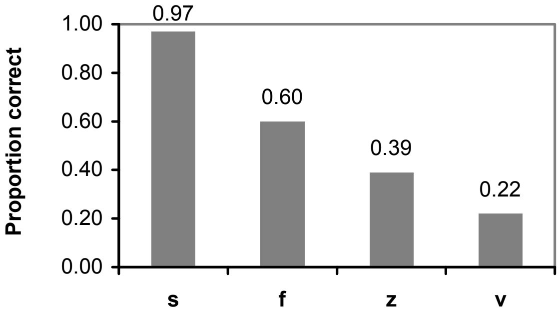
(3) Possible response types

Results: Accuracy. The effect of the first segment was examined with a one-way ANOVA. The independent variable was the first segment (/s/, /f/, /z/ and /v/), collapsing over all individual clusters in the category. The dependent variable was proportion correct. Subjects were treated as a random factor. For statistical tests, each data point corresponds to each participant's proportion correct for each cluster. Thus, for example, if a subject was correct

ISBN 1-876346-48-5 © 2003 UAB

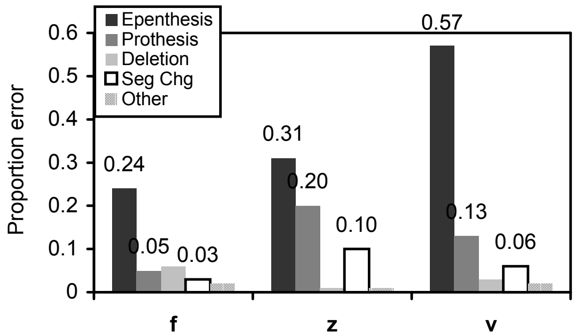
on 3 out of 4 attempts at the cluster /fn/, then she was given a score of .75 for /fn/. This was done in order to have a continuous value for the dependent variable.

Mean proportion correct for each first segment is presented in (4). Results show a main effect of first segment type [F(3,57)=96.78 *p*<.001]. Pairwise comparisons show that each of the initial segment categories are all significantly different from one another (*p*<.001). These results indicate that the nature of the first consonant has a crucial effect on accuracy: speakers produce /f/-initial clusters more accurately than /z/-initial ones, which are more accurate than /v/-initial sequences. As expected, speakers are nearly perfect on /s/-initial sequences.



(4) Performance based on initial fricative

Results: Errors. The distribution of error types for is shown in (5). The single largest error type for all first segments is epenthesis. There is some prothesis, especially corresponding to the production of /z/, consistent with cross-linguistic data showing that using prothesis to repair illegal clusters generally occurs with sibilants [16]. There are also a moderate number of segment change errors for /z/, which becomes the legal cluster-initial /s/.



(5) Error types

4. DISCUSSION

The results of the experiment conform to the prediction made in Possibility 3 in (1): /sC/ ≻ /fC/ ≻ /zC/ ≻ /vC/. These findings suggest that weak-intensity non-sibilant fricatives are disadvantaged relative to /s/, but voice is even more detrimental to English speakers' ability to produce the cluster, as evidenced by the decreased accuracy on /z/ relative to /f/. As expected, /v/-initial clusters induce the poorest performance of all the fricatives, since /v/ is both non-sibilant (weak intensity) and voiced.

The fact that English speakers produce /f/, /z/, and /v/-initial onset clusters with significantly different accuracy is notable. These distinctions are unlikely to result from an intrinsic phonetic relationship between perceptual effects (weak intensity non-sibilant fricatives) and articulatory factors (voiced initial obstruent sequences). That is, although perceptually and articulatorily disadvantaged phonotactic sequences may be dispreferred with respect to those sequences that are salient and/or simple to produce, there is no sense in which either articulatory or perceptual drives must always be maximized above the other. Rather, this is a language-specific decision [17, 18]. Consequently, /f/, /z/, and /v/ can be considered "marked" as in the traditional phonological sense, though this label can be understood as having perceptual and articulatory (i.e. phonetic) origins.

The absence of a phonetically constrained implicational relationship, at least among clusters starting with /f/ or /z/, is supported by cross-linguistic data. For example, while languages like Dutch or Norwegian allow some /f/-initial clusters but no /z/-initial ones, languages like Italian and Serbo-Croatian have /z/-initial clusters but not /f/-initial ones. However, if a language has /v/-initial clusters, like Tsou, Greek, or many Slavic languages, then they also have /f/ and /z/-initial ones (note that all of these languages have the phonemes /s/, /f/, /z/, and /v/).

Given that cross-linguistic cluster inventories exhibit different patterns that allow or prohibit a given phonotactic structure, the relationships among /f/, /z/, and /v/-initial clusters can be said to have become phonologized in the sense of Lindblom [17] or Ohala [18]. In other words, though these fricatives are clearly disadvantaged in clusters on phonetic grounds, making them less likely to be found in legal phonotactic sequences cross-linguistically, they are nevertheless governed by phonological grammars. Crucially, it is the grammar—not just phonetic difficulty—that determines which clusters, if any, are allowed in a given language. It is precisely the fact that both perceptual and articulatory factors must be taken into account that supports a phonological analysis of the results, since phonetic implementation difficulties cannot account for the role of perceptual salience in causing /f/-initial clusters to be poor phonotactic sequences in English.

Although none of these clusters are found in English, the fact that speakers distinguish among them in production indicates that speakers represent and are able to access this kind of phonological knowledge when necessary. Consequently, the results suggest that the English grammar represents /f/, /z/, and /v/-initial clusters in a "hidden" markedness relation, where producing /v/-initial clusters is less optimal than producing /z/-initial ones, and so on, even though this information cannot be gleaned from the lexical items of English. That speakers would have such knowledge is not surprising, considering that it is necessary for dealing with a number of situations, including loanwords and second language acquisition. For example, it has also been argued that hidden phonological knowledge can account for different substitution preferences by speakers of various languages acquiring the English phoneme /θ/ [19].

 ISBN 1-876346-48-5 © 2003 UAB

The existence of hidden markedness relationships among /s/, /f/, /z/, and /v/-initial clusters ultimately leads to the question of their origin. It seems that there are three possibilities: they result from a universal hierarchy; they are a default relationship present in the initial state; or they are language-specific. If the relationships are universal, then all languages should respect the hierarchy found in the experimental results (/sC/ ≻ /fC/ ≻ /zC/ ≻ /vC/), but it has already been noted that there are languages that have /zC/ clusters but not /fC/ ones. If the relationships are present in the initial state, but can be modified based on input from one's native language, then speakers of all languages that do not have /f/, /z/, or /v/-initial clusters should perform like the English speakers on the production task tested in this experiment. Finally, language-specific relationships could perhaps result from differences in how the learning process occurs for distinct languages [20]. In this case, significant cross-linguistic variation on the task examined here would be expected. The experiments needed to answer to this question are an issue for future research.

5. CONCLUSIONS

The results from this study have shown that in a production task, English speakers distinguish among non-native word-initial clusters which are all illegal in English. The non-native clusters used in the experiment—/fC/, /zC/, and /vC/—have unfavorable phonetic characteristics when compared to the legal onset cluster /sC/. Whereas non-sibilant /f/ is less optimal than /s/ for perceptual reasons, /z/ is articulatorily disadvantaged in voiced obstruent clusters. Since /v/ is both non-sibilant and voiced, /vC/ initial clusters contain the least favorable phonetic environment of the experimental clusters. However, the lack of an intrinsic phonetic relationship among these clusters suggests that the distinction among them is encoded at a phonological level, which speakers can access during the course of the experimental task. The representation of phonotactic sequences and the ability of phonological grammars to distinguish among them are independently necessary, since languages can have very different word-initial cluster inventories.

ACKNOWLEDGEMENTS

I would like to thank Paul Smolensky, Matt Goldrick, and audiences at MIT and JHU for comments on this work. This work was supported by an NSF IGERT grant to the Johns Hopkins Department of Cognitive Science.

REFERENCES

[1] E. Broselow and D. Finer, "Parameter Setting in Second Language Phonology and Syntax," *Second Language Research*, vol. 7, pp. 35-59, 1991.

[2] L. Davidson, P. Smolensky, and P. Jusczyk, "The initial and final states: Theoretical implications and experimental explorations of richness of the base," in *Fixing Priorities: Constraints in Phonological Acquisition*, R. Kager, W. Zonneveld, and J. Pater, Eds. Cambridge: Cambridge University Press, in press.

[3] D. Steriade, *Phonetics in Phonology: The Case of Laryngeal Neutralization.* ms., UCLA, 1997.

[4] C. Wilson, "Consonant cluster neutralisation and targeted constraints," *Phonology*, vol. 18, pp. 147-197, 2001.

[5] E. Flemming, *Auditory Representations in Phonology*, Ph.D. dissertation, UCLA, 1995.

[6] R. Kirchner, *An Effort-Based Approach to Consonant Lenition*. London: Routledge, 1998/2001.

[7] K. Stevens, *Acoustic Phonetics*. Cambridge, MA: MIT Press, 1998.

[8] J. Kingston, "Articulatory binding," in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, J. Kingston and M. Beckman, Eds. Cambridge: Cambridge University Press, 1990.

[9] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *Journal of the Acoustical Society of America*, vol. 108, pp. 1252-1263, 2000.

[10] G. Miller and P. Nicely, "An analysis of perceptual confusions among some English consonants," *Journal of the Acoustical Society of America*, vol. 27, pp. 338-352, 1955.

[11] J. Ohala, "Speech aerodynamics," in *The Encyclopedia of Language and Linguistics*, vol. 8, R. E. Asher and J. M. Y. Simpson, Eds. New York: Pergamon Press, 1994, pp. 4144-4148.

[12] K. Stevens, S. Blumstein, L. Glicksman, M. Burton, and K. Kurowski, "Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters," *Journal of the Acoustical Society of America*, vol. 91, pp. 2979-3000, 1992.

[13] J. Ohala and H. Kawasaki-Fukumori, "Alternatives to the sonority hierarchy for explaining segmental sequential constraints," in *Language and Its Ecology*, S. Eliasson and E. H. Jahr, Eds. Berlin: Mouton de Gruyter, 1997.

[14] J. Westbury and P. Keating, "On the naturalness of stop consonant voicing," *Journal of Linguistics*, vol. 22, pp. 145-166, 1986.

[15] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer (v. 4.0.29)*. Computer program, http://www.praat.org, 2002.

[16] H. Fleischhacker, "Cluster-dependent epenthesis assymetries," *UCLA Working Papers in Linguistics*, vol. 7, pp. 71-116, 2001.

[17] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech Production and Speech Modeling*, W. Hardcastle and A. Marchal, Eds. Netherlands: Kluwer, 1990, pp. 403-439.

[18] J. Ohala, "The origin of sound patterns in vocal tract constraints," in *The Production of Speech*, P. MacNeilage, Ed. New York: Springer-Verlag, 1983.

[19] K. O'Connor, *Hidden rankings in Optimality Theory: Evidence from second language acquisition*. Talk presented at the GLOW Workshop on Phonological Acquisition, Utrecht, The Netherlands, 2002.

[20] B. Tesar and P. Smolensky, *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press, 2000.

ISBN 1-876346-48-5 © 2003 UAB