Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Phonetics 34 (2006) 104–137
Journal of Phonetics
www.elsevier.com/locate/phonetics

Phonology, phonetics, or frequency: Influences on the production of non-native sequences

Lisa Davidson*

Department of Linguistics, New York University, 719 Broadway, 4th Floor, New York, NY 10003, USA

Received 10 September 2004; received in revised form 20 March 2005; accepted 30 March 2005

Abstract

This article examines the influence of phonetic and phonological factors and lexical frequency on accuracy and error types in the production of non-native phonotactics. In Experiment 1, English speakers were presented with non-native word-initial consonant clusters that were varied on several phonetic dimensions. Results showed that speakers are not equally accurate on the production of different illegal sequences. An analysis of lexical frequency statistics demonstrates that the frequency of these sequences in other positions across the lexicon does not correlate with accuracy. An explanation based on phonological knowledge is posited instead. A second experiment on the investigation of the strategies used to repair the illegal clusters indicated that speakers prefer schwa insertion. While previous research has assumed that such repairs are vowel epenthesis, a detailed acoustic analysis indicates that inserted schwas are significantly different than lexical schwas. These acoustic characteristics are compatible with articulatory evidence suggesting that there is a prohibition on applying canonical English consonant cluster coordination to phonotactically illegal sequences, leading speakers to "pull apart" the consonant gestures and causing a transitional schwa to appear on the acoustic record. The ramifications of these results for the role of an abstract phonological level in production are discussed.

*Tel.: +1 212 992 8761.
E-mail address: lisa.davidson@nyu.edu.

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2005.03.004

1. Introduction

Researchers agree that in the production of real words, there are a number of factors influencing what constitutes a well-formed word of a language. On one hand, there are some types of sequences that are prohibited outright and are not attested in the lexicon at all. Particular segments may be completely unattested in certain positions even though the segment itself is part of the inventory of the language (e.g., /h/ in English: *home*, but *[moh]; cf. [tuh] 'rain' in Zoque (Wonderly, 1951)). It is commonly assumed that since phonotactic restrictions are language-particular, they are encoded in the higher-level phonological grammar of the language.

On the other hand, phonological well-formedness is not a simple, categorical concept. That is, a sequence type may be attested but infrequent in the lexicon, which has a number of consequences for the processing of such sequences. That speakers are sensitive to the frequency of different types of elements in the lexicon has been demonstrated by a growing body of research. Frequency effects have been found in both production and perception tasks. For example, when asked to repeat non-words, speakers respond more rapidly when the items contain high-probability phoneme sequences (e.g., Vitevich & Luce, 1999, 2005; Vitevich, Luce, Charles-Luce, & Kemmerer, 1997). Perception studies have shown that when presented with acoustically ambiguous consonants in a sequence, listeners are more likely to interpret them as the consonant that would comprise the more frequent sequence (Pitt & McQueen, 1998). Likewise, in a word recognition task, participants are more likely to recall non-words with high-probability sequences than with low-probability sequence (Frisch, Large, & Pisoni, 2000). It has also been shown that acquisition of lexical items is facilitated when words have a greater phonotactic probability (Storkel, 2001; Storkel & Rogers, 2000). Finally, a number of studies have shown that participants asked to rate the "wordlikeness" of non-word stimuli generally give higher ratings to stimuli containing sequences that are attested more frequently across the lexicon (Bailey & Hahn, 2001; Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Munson, 2001; Vitevich et al., 1997).

While many previous studies have focused on the production of non-words as a means to test how frequency of a sequence affects phonological processing (e.g., Edwards, Beckman, & Munson, 2004; Hay, Pierrehumbert, & Beckman, 2001; Munson, 2001), none have focused solely on how frequency may affect the processing of sequences that are not attested at all in certain positions. Non-attested sequences refers not to sequences containing phonemes not found in a language (e.g., [xt] for English), but rather to sequences of phonemes that are positionally restricted (e.g., *[ft] or *[pk] word-initially in English, but cf. *after* or *napkin*). Though speakers have no experience with such sequences word-initially, it is possible that distributional information calculated across other positions in lexical items may affect speakers' production of unattested sequences. If this is true, it would have important implications for factors affecting second language acquisition and the phonology of loanwords.

Alternatively, it may be that distributional and frequency information from other positions in the lexicon do not affect the production of non-native sequences. The alternative is that the production of such sequences is determined by phonological factors related to the speakers' native language. Relevant phonological influences are discussed in the following section.

1.1. Production and perception of non-native sequences

The second language acquisition literature provides ample evidence suggesting that speakers do not produce all unattested sequences with equal accuracy. For example, an investigation of the production of English codas by Vietnamese speakers showed that while speakers had moderate trouble producing /s/ and /f/ in coda position, they were less accurate on /v/ and /l/, followed by /ʃ/ (Hansen, 2004). All of these are phonemes of Vietnamese, though they are not allowed in coda position. Broselow and Finer (1991) and Eckman and Iverson (1993) found that in producing English initial consonant clusters that are phonotactically prohibited in their native languages, Japanese and Korean participants were more accurate on voiceless stop-initial sequences than they were on either voiced stop or fricative-initial ones. In a study investigating the role of sonority sequencing in the production of phonotactically illegal sequences, Davidson, Jusczyk and Smolensky (2004) presented English speakers with words containing Polish-legal onset clusters such as /kt/, /tf/, /dv/, /zm/, /vn/ etc. Results showed that sonority sequencing alone cannot account for the data. For example, clusters like /zm/ and /vn/ differ only in place, not in sonority distance, but speakers' accuracy in producing these sequences is very different (63% vs. 11%). Similar accuracy results were reported for Scottish English speakers producing Russian onset clusters who were asked to write down what they heard (Haunz, 2003).

These studies focusing on the production of non-native phonotactics have tended to rely on grammatical concepts like markedness to explain their findings. For example, in accounting for his finding that Spanish speakers are more accurate on English two-member initial /s/-clusters (e.g., *spat*) than three-member clusters (e.g., *splat*), Carlisle (1998) notes that two-member sequences are less marked than three-member ones. Broselow and Finer (1991) also rely on markedness to explain their findings; for example, they argue that Japanese and Korean speakers performed equally well on relatively unmarked but unattested clusters such as [pr] as they did on more marked clusters such as [bj], which are allowed in the language. In this case, [pr] is less marked than [bj] since it contains a voiceless initial stop. In their Optimality Theoretic analysis of their findings that Mandarin Chinese speakers devoice voiced obstruents in syllable final position, Broselow, Chen and Wang (1998) argue that even second language speech exhibits the phenomenon known as "emergence of the unmarked" (McCarthy & Prince, 1994). That is, though neither type of final obstruents is permitted in Mandarin, speakers acquiring a language will exhibit the unmarked instantiation of a given segment before producing the marked version.

Perception studies have also focused on the processing of non-native sequences to examine the role of the native language as a phonological filter. Recent studies of on- and off-line perceptual processing of illegal clusters have shown that listeners tend to assimilate impossible clusters to sequences that are legal in their native languages. Based on results from a transcription task, a forced choice paradigm, a phonemic gating task, and a phoneme monitoring task, Hallé, Segui, Frauenfelder and Meunier (1998) found that French listeners are more likely to hear /tl/ and /dl/ word-initial sequences as /kl/ and /gl/. Dupoux and colleagues demonstrated that Japanese listeners faced with consonant sequences that are prohibited by the native phonology, such as *ebzo*, perceive an illusory epenthetic [u] between the consonants, and cannot distinguish among stimuli without the vowel (*ebzo*) and those with one (*ebuzo*) (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; see also Kabak, 2003 and Massaro & Cohen, 1983). These authors conclude that higher-level knowledge of phonotactics must ultimately dominate lower level perceptual processes which can otherwise effortlessly distinguish native segments in legal contexts.

Pitt (1998) showed that in the perception of illegal clusters in English, listeners are more sensitive to absolute phonotactic constraints than they are to frequency considerations. Pitt hypothesized that if frequency is more influential than phonotactics in recognizing consonant clusters, then listeners would be more biased toward hearing the more frequent of pairs like /br/-/bl/ or /tr/-/tl/. Frequency calculations were made across word-initial and word-medial clusters, so clusters like /tl/ (as in *atlas*) were also assigned a frequency. Results showed that the correct liquid was identified for all legal sequences and that there was no bias toward reporting the more frequent of the cluster types. For the sequences that are illegal in some positions, there was a strong bias toward hearing a legal sequence, without any interference from frequency.

Moreton (2002) designed a study to investigate whether structural models (those which explain perceptual processing in terms of phonological generalizations over classes of phonemes) or unit models (those which are concerned with frequency, lexicality, etc.) can better account for perceptual data. Moreton contrasted English listeners' performance on nonsense words beginning with either /bw/ or /dl/ clusters. The initial sequences /bw/ and /dl/ both have zero frequency in the English lexicon, but structurally, English contains onset consonants with the phonological properties similar to /bw/, but not /dl/. For example, Moreton hypothesizes that /bw/ is influenced by the existence of /br/: /r/ has secondary rounding, so legal sequences like /br/ indicate that [labial][labial] sequences are allowed. Results show that speakers were significantly more likely to correctly perceive /bw/ than /dl/. Moreton argues that the bias in favor of /bw/ can only be accounted for in terms of structural models that incorporate higher level phonological knowledge. Findings similar to those of Moreton were also reported by Coetzee (2004), who found that degraded perception of words violating the Obligatory Contour Principle (e.g., [spʌp], [skɛk]) could not be reduced to lexical statistical effects, and Kabak (2003), who argued that perception of word-medial consonant sequences by Korean speakers was sensitive to whether or not the first consonant was a possible coda or not, and not to any characteristics of the lexicon.

The aim of the first experiment in this paper is to contribute to the literature comparing the role of frequency versus the role of structural or phonological knowledge in the production of unattested sequences. In Experiment 1, speakers are presented with words containing /f/-initial, /z/-initial, and /v/-initial sequences, all of which are unattested in English words-initially. Four different types of frequency measures are calculated over other positions in the word for each of the non-native sequences presented, and speakers' accuracy on the unattested sequences is evaluated with respect to the frequency measures.

1.2. Errors in the production of non-native sequences

The production of non-native sequences not only raises the question of which sequences are produced with the lowest accuracy, but also the issue of how speakers ultimately repair them. Thus, the second experiment focuses on the nature of the repairs. While speakers may choose from a variety of strategies, such as deletion, epenthesis, segment change, etc, previous research has demonstrated that they most often repair non-native consonant clusters by inserting a vowel between the two consonants (Davidson et al., 2004; Hancin-Bhatt & Bhatt, 1998; Tarone, 1987). Earlier studies have generally assumed that insertion repairs reflect the epenthesis of a lexical vowel, often schwa. However, results emerging from research within Articulatory Phonology have raised the possibility that the vowel present between the consonants in the unattested sequences is not the insertion of a lexical vowel, but rather the failure to adequately overlap the consonant gestures (Browman & Goldstein, 1990; Gafos, 2002; Jannedy, 1994). This possibility can be examined acoustically by demonstrating that the inserted vowel does not have the same phonetic properties as a lexical schwa. If such a difference is found, there are two possible reasons that would lead to distinctions between lexical and inserted schwa.

The first possibility is that the articulatory coordination of adjacent segments influences the nature of repairs. It has been noted that temporal dimensions such as phoneme duration, vowel reduction, voice onset time, and decreased overlap leading to releases between consonant segments are all factors that can lead to the perception of a foreign accent in a non-native speaker (Cebrian, 2000; Flege, Munro, & MacKay, 1995; Magen, 1998; Major, 1987; Tajima, Port, & Dalby, 1997; Zsiga, 2003). Speakers attempting to produce non-native sequences firstly must discover when the combination of certain adjacent consonants is permitted, either in specific positions or across-the-board; and secondly, even when speakers learn that particular sequences should be treated as a unit, they must still discover the articulatory patterns required to properly coordinate the segments of the string. Some recent articulatory studies suggest that epenthetic vowels used to repair phonotactically prohibited sequences may not result from vowel insertion, but from the failure to produce the consonants using the correct overlapping gestural coordination for the target language (Davidson, 2005; Smorodinsky, 2002). That is, if speakers produce the consonants with an insufficiently overlapping configuration ("mistiming"), a transitional vocoid would result between the segments in the clusters (Bradley, 2004; Gafos, 2002; Hall, 2003).

The other possibility is that speakers are inserting a schwa, but in an effort to produce the sequence as close as possible to the native speaker stimulus, they are reducing the duration of the vowel and thus failing to reach the target formant frequencies of the vowel. This phenomenon has previously been called undershoot (Lindblom, 1963). Lindblom demonstrated for Swedish that as vowels decreased in duration, they often failed to reach their target formant frequencies. He argued that the directionality of the formant movement is not consistent with centralization or reduction of the vowel to schwa; rather, speakers cannot achieve the canonical target of the vowel. If this is happening when speakers repair non-native sequences with schwa insertion, predictions can be made about the duration and expected changes in formant values for the epenthesized schwa. These are discussed in the introduction to Experiment 2.

The remainder of this paper is as follows. The first experiment is a study of English speakers' production of fricative-initial consonant clusters in word-initial position. In the stimuli, the place and voice of the fricative and the manner of the second consonant were varied. The frequency of the combinations of fricatives and following obstruents or nasals in other positions in English words was compared to production accuracy. In the second experiment, speakers were asked to produce stimuli both with consonant clusters and with a schwa between the consonants in order to examine whether the acoustic properties of the inserted vowel are the same as those for words with underlying schwa. In the last section, the factors influencing production both with respect to the accuracy results and the repair results are discussed.

2. Experiment 1

2.1. Participants

The participants were 20 Johns Hopkins University undergraduates who received course credit for their participation. All of them were native speakers of English and had no exposure to Slavic languages. None reported any history of speech or hearing impairments.

2.2. Materials

The target words used in the study were pseudo-Czech words with /s/, /f/, /z/, and /v/-initial obstruent clusters. These initial segments were combined with the stops /p/, /t/, /k/ for the voiceless fricatives and /b/, /d/, /g/ for the voiced fricatives, the other fricative with the same voicing specification (i.e., /sf/, /fs/, /zv/ and /vz/), and the nasals /m/ and /n/ to create 24 word-initial clusters. All possible combinations are given in Table 1. 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. These words are shown in the appendix.

For each of the experimental clusters, four different frequency measures were calculated across the English lexicon. Using the online MRC Database (Wilson, 1988), both type and token frequencies per million words were calculated for morphologically simplex and morphologically complex words. The frequency measures used in this experiment are similar to those in Hay et al. (2001), except that instead of using log probability, the log of the cluster frequency is used. Since the clusters in question are not attested in initial position, it was decided that it is not the transitional probability of the phonemes in the cluster that is important, but how many type and token occurrences of the cluster exist in the lexicon. The raw cluster frequency is based on the

Table 1
Pseudo-Czech word-initial clusters used in experiment

C1	C2	Cluster	Example
/s/	nasal	/sm/,/sn/	smava
	fricative	/sf/	sfano
	Stop	/sp/,/st/,/sk/	stamo
/f/	nasal	/fm/,/fn/	fnada
	fricative	/fs/	fsapi
	stop	/fp/,/ft/,/fk/	fkale
/z/	nasal	/zm/,/zn/	znafe
	fricative	/zv/	zvabu
	stop	/zb/,/zd/,/zg/	zdaba
/v/	nasal	/vm/,/vn/	vmala
	fricative	/vz/	vzamo
	stop	/vb/,/vd/,/vg/	vbaki

Table 2
Log frequency of occurrence of the 18 target word-initial clusters in word-medial and final position, calculated from the Kucera and Francis written frequency counts of the MRC Database

Cluster	Type frequency (per million)		Token frequency (per million)	
	Monomorphemic	Multimorphemic	Monomorphemic	Multimorphemic
fm	0	0	0	0
fn	0	2.20	0	2.20
fs	ln(0)	5.13	ln(0)	5.21
fp	ln(0)	ln(0)	ln(0)	ln(0)
ft	3.97	5.71	7.64	7.86
fk	ln(0)	0	ln(0)	0
vm	ln(0)	1.95	ln(0)	5.58
vn	ln(0)	3.47	ln(0)	3.47
vz	0	5.58	0	6.54
vb	ln(0)	0	ln(0)	ln(0)
vd	ln(0)	4.71	ln(0)	6.33
vg	ln(0)	ln(0)	ln(0)	ln(0)
zm	3.09	5.05	4.38	5.62
zn	0	1.10	5.97	5.98
zv	ln(0)	ln(0)	ln(0)	ln(0)
zb	2.64	3.33	5.04	5.13
zd	1.95	6.02	5.16	7.37
zg	ln(0)	ln(0)	ln(0)	ln(0)

Cells labeled '0' have a raw frequency of 1, and cells labeled 'ln(0)' have a raw frequency of 0.

number of words per million containing the cluster. The clusters were included in the frequency counts regardless of where they occurred in the word, encompassing both word-medial and word-final position (e.g., /ft/ in *daft*, *after*). Thus, this method ignores potential differences in syllable position. All words are real, including place names recognized in American English (e.g., /zb/ in *Lisbon*). The frequency counts are shown in Table 2. Since some clusters are not attested in any position, the log cannot be calculated for these, and their log frequency is represented in the table as ln(0).

Note that in morphologically complex words, especially those ending in /z/ or /d/, the relevant cluster is often located at the morpheme boundary (e.g., /vd/ in *loved*, /vz/ in *lives*). These forms are included because they provide insight into the knowledge that speakers may recruit for producing non-native sequences. Specifically, it could be that speakers are relying on their articulatory knowledge from other parts of the word to produce the initial sequences. Though studies have often based their frequencies counts on monomorphemic lexical items (e.g., Hay et al., 2001; Munson, 2001), in this study it is also important to include morphologically complex words to determine whether or not speakers may use the articulatory programs they have developed for producing the relevant sequences in morphologically complex forms.

Table 3
Possible response types

Response type	Definition	Example
Correct	Target is produced with no changes or simplifications	/zvabu/ → [zvabu]
Insertion	Target is produced with a schwa between the consonants in the cluster	/zvabu/ → [zəvabu]
Deletion	Target is produced with either the first or second member deleted	/zvabu/ → [zabu] /zvabu/ → [vabu]
Prothesis	Target is produced with a schwa before the cluster	/zvabu/ → [əzvabu]
Segment change	Target is produced with two segments, but one differs from the original	/zvabu/ → [svabu]
Other	Target is not produced, has more than one error, or is completely unrecognizable	/zvabu/ → ∅ /zvabu/ → [vəvabu] /zvabu/ → [spada]

2.3. Design and procedure

The experiment was designed with PsyScope 1.2.6 and was presented on a Macintosh G3 laptop. Participants were seated in a sound-proof booth and their responses were digitally recorded with a CSL using a head-mounted, unidirectional short-range microphone.

All 96 words were presented in random order to all participants in a repetition task. Results from similar previous testing indicate that speakers perform essentially the same regardless of whether they simply repeat the target words or produce them embedded in an English sentence (see Davidson, 2003; Davidson et al., 2004). The repetition task was chosen for its greater simplicity. 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.[1] Twenty milliseconds after the word appeared on the screen, the target stimulus was presented auditorily to the participant through external speakers. The word was repeated again 300 ms 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 2500 ms. Participants were given five practice trials before the experimental trials.

In order to determine the participants' responses, the waveform and spectrogram for each target word were analyzed using Praat (Boersma & Weenink, 2004) to determine what, if any, error had been made. There were six possible response categories for each target: Correct (C), Insertion (I), Prothesis (P), Deletion (D), Segment Change (S), and Other (O). All responses were coded by the author, and 15% of the tokens were additionally coded by a second transcriber blind to the purpose of the study. Interrater reliability was 91%.

A token was coded for insertion under the following circumstances: for /C/ + nasal sequences, a period of voicing after frication with formant structure containing a visible second formant that

[1]The target words were presented orthographically in order to bias the participant toward hearing the word-initial cluster. While the use of orthography may present a confound, it was felt that presenting the participant with the written form of the target was critical for preventing possible misperceptions of the word.

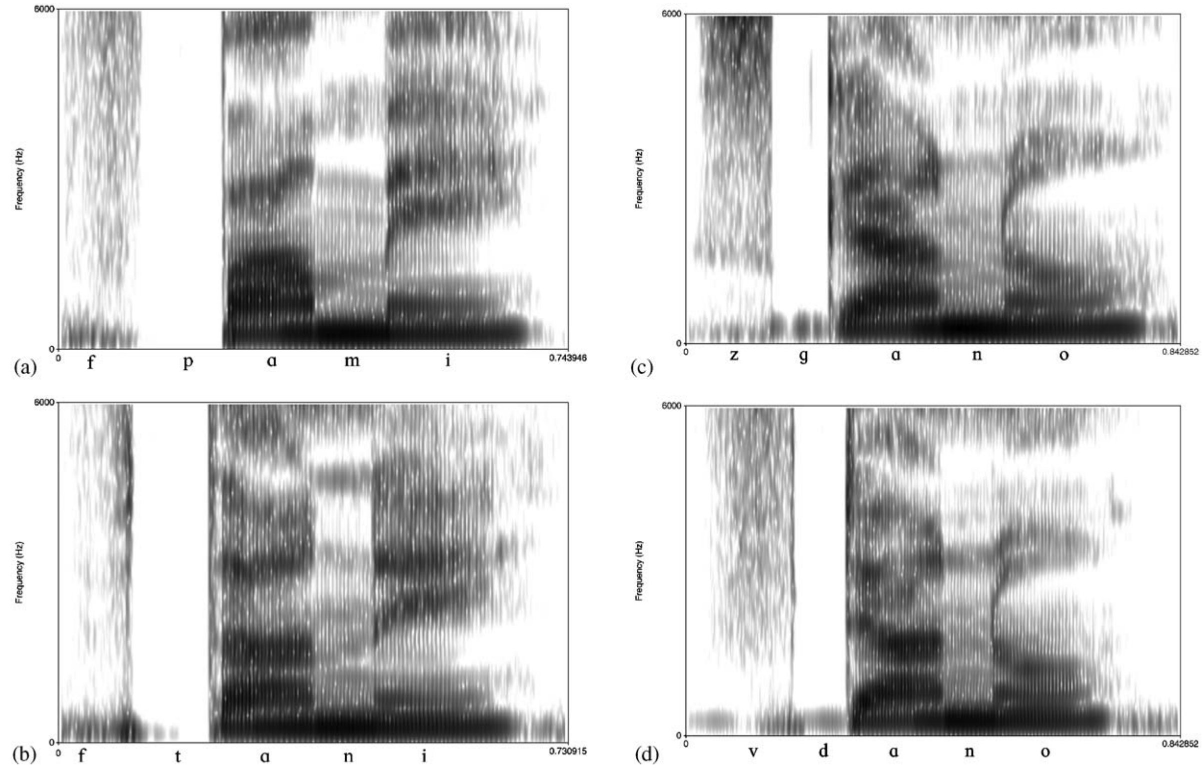

Fig. 1. Spectrograms of various stimuli produced by a Czech speaker. (a) A voiceless cluster with uniform frication in the word [fpami]. (b) A voiceless cluster with a burst-like release in the word [ftani]. (c) A voiced cluster with a uniform frication in the word [zgano]. (d) A voiced cluster with a burst-like release in the word [vdano]. Note that neither (b) nor (d) contain a vowel between the consonants.

ended with abrupt lowering of intensity at the onset of the nasal was present; /f/ + obstruent contained any period of voicing between the two consonants (including both formants and voicebars); and /z/ + obstruent and /v/ + obstruent contained a second formant that ceased before the closure or frication of the next obstruent. The remaining repairs are exemplified in Table 3.

In the Czech production of both voiced and voiceless clusters, there is no vocalic material present in the inter-consonantal position. For /f/-initial sequences, the Czech speaker generally produced the fricative with unvarying intensity, with the spectral energy uniformly distributed over the spectrum. This is demonstrated in Fig. 1a. In a very small number of /f/ + stop clusters, there was a burst-like increase in intensity of the frication at the very end of the /f/, which may be indicative of the moment of release in these tokens. However, there was no evidence of vocalic material following the fricatives in these tokens. This is shown in Fig. 1b. Like the voiceless stimuli, the voiced tokens usually also display uniform frication intensity throughout the duration of the fricative, though occasionally voiced fricative + stop clusters also show evidence of a burst-like release (Fig. 1c and d). Again, no vowel is present between the consonants.

Given these characteristics of the stimuli, English speakers' responses were only coded for correct if they had the same characteristics as the Czech stimuli. In cases where English speakers

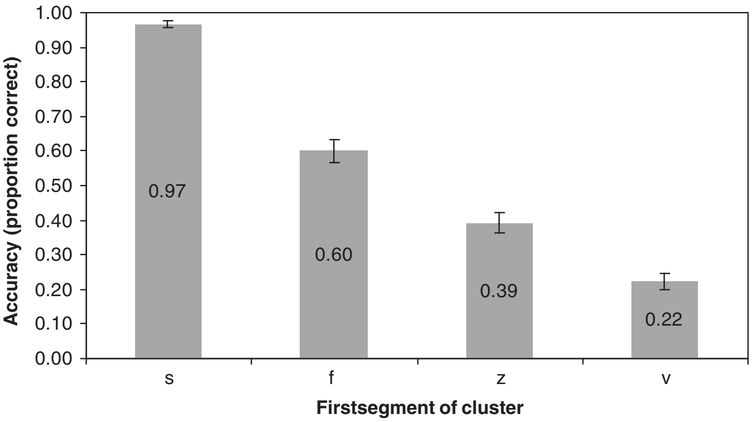

Fig. 2. Proportion correct on each cluster type plotted by first segment (Experiment 1). Error bars indicate standard error.

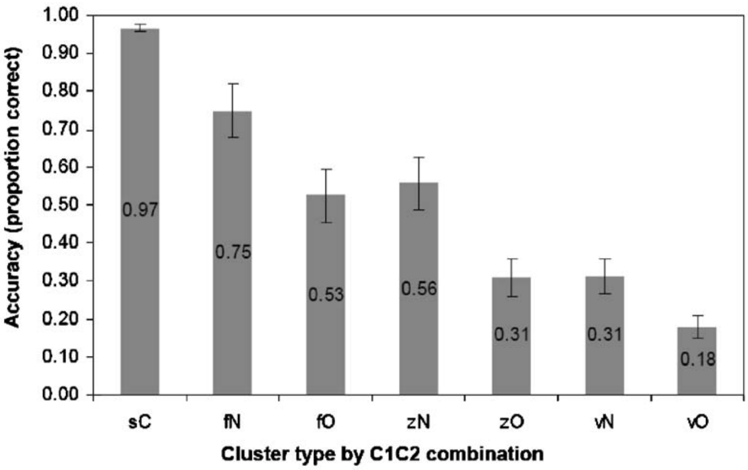

Fig. 3. Proportion correct on clusters broken down by cluster type (N = nasal, O = obstruent) (Experiment 1). Error bars indicate standard error.

showed evidence of a release but no following vocalic material, the production was coded as correct. Likewise, tokens were coded for correct if the clusters produced by the participants matched the manner, place, and voice specifications of the input, and the consonants were produced in the correct linear order. Other slight variations from the target stimulus, such as duration, did not prevent the token from being coded as correct.

2.4. Results

2.4.1. Effect of first and second consonant

The effect of first segment was examined with a univariate analysis of variance (ANOVA). The independent variables were the first segment (/s/, /f/, /z/, /v/) and the second segment (obstruent vs. nasal). The dependent variable was the average proportion of correct responses for each cluster type. This proportion was arcsine transformed before being submitted to the ANOVA. Participants in this and all following statistical procedures were treated as a random factor. Mean proportion correct for each first segment is presented in Fig. 2. Results show a main effect of first segment type [$F(3,57) = 90.16$, $p < .001$, partial $\eta^2 = .826$] and a main effect of second segment type [$F(1,19) = 27.07$, $p < .001$, partial $\eta^2 = .588$]. Scheffé post hoc comparisons show that each of the initial segment categories are all significantly different from one another ($p < .001$). The interaction between first and second segment type is also significant [$F(3,57) = 4.30$, $p < .008$, partial $\eta^2 = .185$]. A graph of the proportion correct broken down by C1C2 combination which illustrates the interaction between C1 and C2 is shown in Fig. 3. Scheffé post hoc tests comparing each combination of C1 and C2 gives the following groupings: sC $\succ$ fN $\succ$ fO, zN $\succ$ zO, vN $\succ$ vO ($p < .001$). The symbol '$\succ$' between a pair of clusters indicates that they are significantly different according to the post hoc test.

These results indicate that the nature of both the first and second consonants 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. Simultaneously considering both the first and second consonants indicates that speakers are sensitive not only to the characteristics of the consonant individually, but also to the relationship that they have with one another. For each first consonant, the speakers are more accurate when the second consonant is a nasal than when it is an obstruent.

2.4.2. *Relationship to frequency*

In order to determine whether the participants' accuracy on /f/, /z/ and /v/-initial sequences can be derived from the frequency of these sequences in other positions across the lexicon, first the accuracy for each individual cluster type averaged over all speakers was computed. The accuracy for each individual cluster averaged over participants was then correlated with the log frequencies from Table 2. Accuracy on the individual non-native sequences is not significantly correlated with any of the frequency measures (monomorphemic type: $r^2 = .14$, $p = .57$; multimorphemic type: $r^2 = -.14$, $p = .59$; monomorphemic token: $r^2 = .14$, $p = .58$; multimorphemic token: $r^2 = -.09$, $p = .72$).

These findings indicate that the four frequency measures examined here cannot explain performance on the individual cluster types. Whereas a significant effect would demonstrate a positive correlation between accuracy and frequency (i.e., higher frequency leads to increased accuracy), the measures are either close to no correlation at all (monomorphemic type and token) or are actually weakly negatively correlated (multimorphemic type and token). While it is possible that a more complex measure of lexical statistics could explain performance on non-attested clusters, those commonly reported in the literature—which are directly related to how often a phonotactic structure can be found across the lexicon—are unlikely to account for the behavior in Experiment 1. Consequently, another explanation will be explored in Section 2.5.1.

2.4.3. *Error types*

The distribution of error types is shown in Fig. 4. In this graph, the errors are averaged over all participants. The single largest error type for all categories of first segments is insertion. There is also some prothesis, especially corresponding to the production of /z/. This repair is consistent

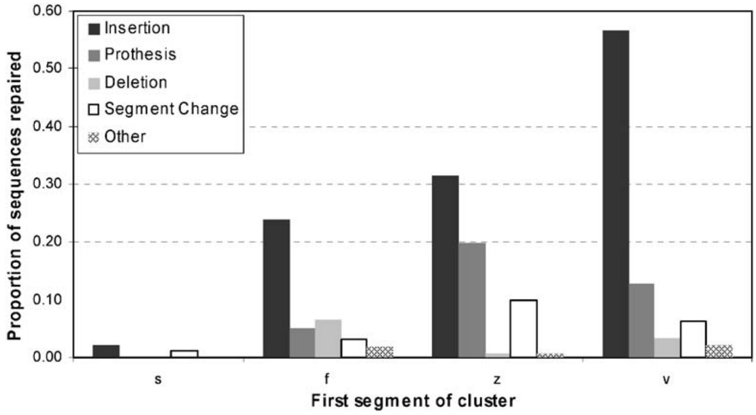

Fig. 4. Proportions of error types for each first segment type, averaged over all participants.

with cross-linguistic data showing that using prothesis to repair illegal clusters generally only occurs with sibilants (Broselow, 1983, 1991; Fleischhacker, 2001). There is also a moderate number of errors in the Segment Change category for /z/, which becomes the legal cluster-initial /s/. An examination of individual participants shows that they all conform to this general pattern, except for one participant, who exhibits only prothesis and no insertion (prothesis for /fC/: 42%; /zC/: 25%, /vC/: 58% of that participant's total errors).

2.5. Discussion

2.5.1. Fricative-initial obstruent clusters

The results of this experiment demonstrate that performance on non-native word-initial clusters is affected by both the first segment of the cluster and the second segment of the cluster. More specifically, the results demonstrate that speakers show the general pattern of accuracy of /sC/ > /fC/ > /zC/ > /vC/.

The lack of correlation with both type and token frequencies of monomorphemic and multimorphemic words shows that there is no relationship between the frequency of the cluster over other positions in the lexicon and accuracy on the individual clusters. An alternative explanation may be sought in phonological (or structural) knowledge (Moreton, 2002). Higher-level phonological knowledge, as discussed by both Moreton and by proponents of Exemplar Theory such as Pierrehumbert (2003), is not concerned with specific combinations of phonemes per se, but with the effect of generalizations that are derived over classes of phonemes or sequences. Though /fC/, /zC/ and /vC/ clusters are not permitted in English, they are phonologically related to existing sequences. For example, English does permit voiceless fricatives as the first member of a word-initial cluster in various cases. Obvious instances of such consonant clusters include /s/-initial sequences (e.g., /st, sm, sl/), /ʃ/-initial sequences (e.g., /ʃr/), and /f/-initial sequences (e.g., /fl, fr/) (Hammond, 1999).

These sequences indicate that English does not disallow word-initial sequences starting with voiceless fricatives, and more importantly, that /f/-initial sequences are permitted in English. These facts may give speakers an advantage in producing word-initial /f/ + obstruent or /f/ + nasal sequences. Word-initial /f/ + obstruent clusters are not necessarily judged by English speakers to be illegal, by analogy to the existence of fricatives (especially /s/ + obstruent, /s/ + nasal sequences) and /f/ in initial clusters.

A similar explanation could underlie the relative level of accuracy for /z/-initial sequences. First, English does not allow any voiced fricatives in word-initial clusters, which would account for the speakers' performance on /z/ and /v/-initial sequences relative to /f/. Yet, speakers are still more accurate on /z/-initial sequences than on /v/-initial sequences. This may be due to the fact that the voiced /z/-initial sequences are very similar to a type of clusters that are permitted initially in English, namely /s/ + obstruent and /s/ + nasal clusters. Speakers often produce coronal fricative-initial clusters, which may facilitate their attempts to produce voiced coronal fricative-initial obstruent and nasal clusters. The /v/-initial clusters, on the other hand, do not have the same direct connection. Whereas /z/ followed by an obstruent or nasal differs from existing sequences primarily on the basis of voicing, /v/ is even further disadvantaged by the fact that /f/ is only found in /f/ + sonorant sequences, not /f/ + obstruent clusters. While speakers may be able to extend their ability to produce /f/ + obstruent clusters by analogy to other existing voiceless initial fricatives (/s/) that combine with nasals and obstruents and /f/ + sonorant sequences, /v/ does not get the same direct benefit, especially relative to /z/-initial sequences.

Though phonological models like those discussed by Moreton (2002) are usually discussed in terms of abstract phonological generalizations built up over existing phonological classes, the preceding explanation of the relationship between success in producing /f/, /z/, and /v/-initial sequences and permitted English sequences may also reflect the ability of English speakers to recruit and modify the articulatory patterns developed for the production of existing sequences. This is not to say, however, that articulatory knowledge is necessarily more peripheral or low-level than other types of phonological knowledge. Recent research, particularly in the Articulatory Phonology framework, has argued that such articulatory patterns are an integral part of phonological generalizations (Browman & Goldstein, 1992a; Byrd, 1996b; Gafos, 2002; Levelt & Wheeldon, 1994; Ussishkin & Wedel, 2003). If this is true, then when speakers do not accurately produce a non-native sequence, is it possible that the resulting errorful production is due to improper coordination of the articulatory gestures comprising the consonants? This issue is raised in the next section addressing speakers' repair strategies.

2.5.2. *Errors*

When producing the non-native consonant clusters, participants repaired them more often with vowel insertion than with any other type of repair. This strategy may be a result of the fact that the stimuli were presented both aurally and orthographically. However, this is not a drawback of the design; instead, it suggests that when speakers are aware of the intended underlying phonemes of the word they are trying to produce, they prefer to preserve as much information about the string as they can (e.g., Abrahamsson, 2003; Weinberger, 1994).

Most studies that report vowel insertion as a repair for phonotactically illegal sequences in non-native language production have assumed that speakers are phonologically epenthesizing a vowel between the consonants of the illegal sequence (e.g., Anderson, 1987; Benus, Smorodinsky, & Gafos, 2004; Broselow, 1983; Broselow & Finer, 1991; Davidson et al., 2004; Eckman & Iverson, 1993; Hancin-Bhatt & Bhatt, 1998). However, it has also been argued that the presence of a vowel, especially a schwa, in the acoustic record does not necessarily arise through epenthesis or deletion

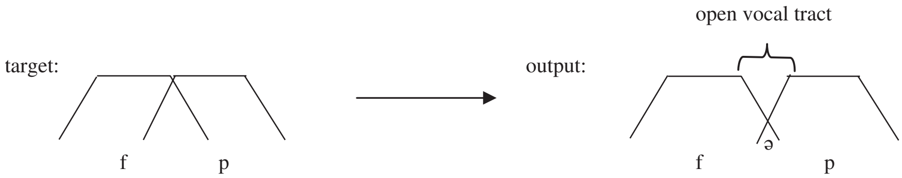

Fig. 5. Transitional schwa between consonants (no gesture corresponding to /ə/). Each consonant is represented by a plateau. The beginning of the horizontal line of the plateau represents the achievement of the articulatory target, and the end of the plateau is the articulatory release. When the release of C1 is not overlapped by the target of C2, a transitional schwa is produced.

of a gesture, but may be a result of lengthening a gesture or changing the timing among two gestures (Browman & Goldstein, 1992b; Jannedy, 1994; Price, 1980). Thus, with respect to the prevalence of insertion used by the speakers in this experiment, one possibility regarding the nature of the insertion repair is that speakers are not actually epenthesizing a vowel, but rather are ‘‘mistiming’’ articulatory gestures with respect to their expected coordination patterns.

This interpretation follows from the idea that speakers may not consistently apply the necessary articulatory patterns for accurately producing non-native sequences with close coordination. In close coordination, there is no release between consonants in a sequence (Catford, 1977; Henderson & Repp, 1982). Given that the participants in the experiment were instructed to produce the clusters correctly, it is possible that the instructions may have precluded production of a schwa, since modifying a cluster with a schwa is a strategy for repairing a prohibited phonotactic sequence, not for producing it correctly.

With respect to the experimental classification as ‘‘insertion’’, a schwa could be produced through phonological epenthesis, but also through a gestural coordination in which the release of the first consonant is not overlapped by the target of the second consonant. Such a strategy has been argued to be the norm for languages such as Moroccan Arabic, which exhibit transitional schwas between consonants in the coda (Gafos, 2002; Hall, 2003). That is, if the consonantal gestures comprising the cluster are produced sufficiently far apart and the speaker is in the speech-ready state, then the vocal tract may be unobstructed long enough for vocal fold adduction and vibration to occur. This would lead to the percept of voicing between the two adjacent consonants.[2] This possibility was tested by Gafos (2002) using the GEST computational model of gestures and gestural dynamics developed at Haskins Laboratories. The simulation showed that a vocalic release was in fact produced for a coordination pattern in which the release of the first consonant was not overlapped by the target of the second consonant. This is exemplified with a schematic demonstrating a transitional schwa in Fig. 5.

The possibility that the schwa arises from a non-overlapping gestural coordination is supported by a closer examination of details of the acoustic record. Spectrographic evidence indicates that there is a considerable amount of variability regarding the duration, intensity and amount of formant structure corresponding to the schwa produced between the two consonants of the non-native cluster. This is the case for both voiceless and voiced clusters. When categorizing voiceless

[2]In general, Browman and Goldstein assume that the default glottal state is the one that produces voicing (Browman & Goldstein, 1986), so if the constrictions and glottal abduction gestures of the consonants are relaxed for long enough, then voicing can occur.

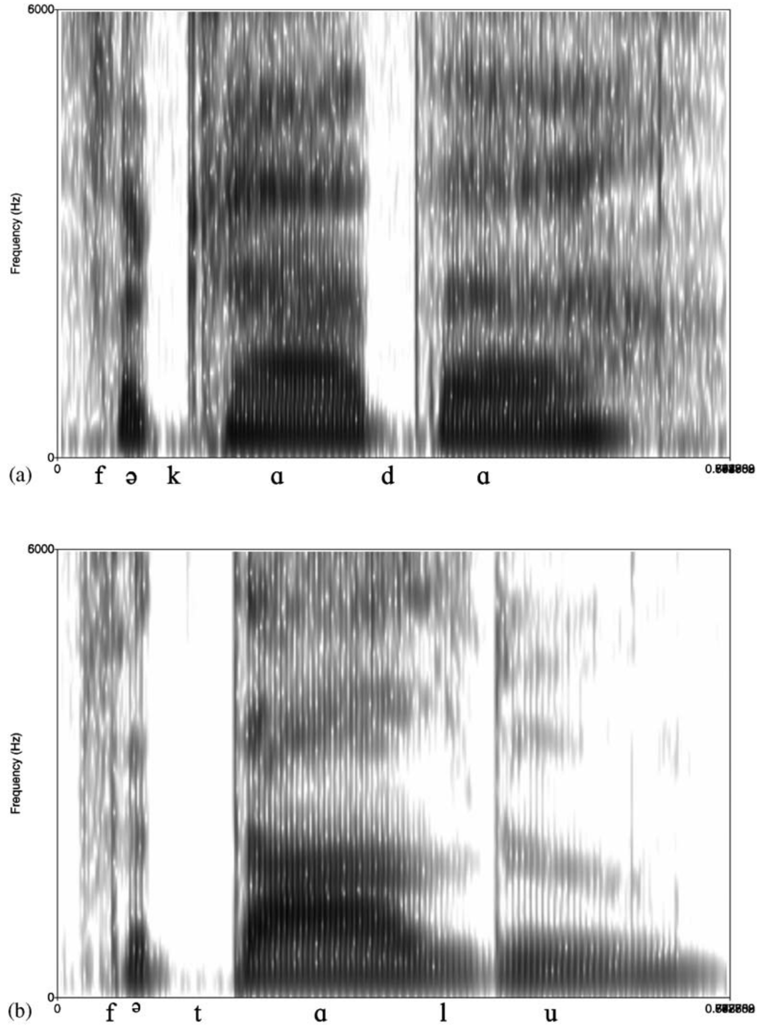

Fig. 6. Spectrograms of different manifestations of inserted vowel between voiceless consonants. The top panel contains a large, robust period of voicing in interconsonantal position in the production of the word /fkada/ → [fəkada] (subject 20). In the bottom panel, the cluster contains a shorter, transitional period of voicing in the word /ftalu/ → [f$^{\text{ə}}$talu] (subject 7). The low-intensity energy present in the voiceless consonants is due to the hum of a laptop computer.

clusters in terms of errors produced by the speakers, an utterance was coded as containing an inserted vowel if any period of voicing intervened between the consonants, whether it was more similar to a lexical schwa, or a short, transitional one. Examples of these cases are shown in Figs. 6 and 7.

That there is so much variability in how the schwa is realized in the English speakers' productions further suggests that speakers are not necessarily epenthesizing a phonological schwa in order to repair the phonotactically ill-formed cluster. The speech rate remained relatively constant during the course of the experiment, so there is no simple explanation for why the schwa

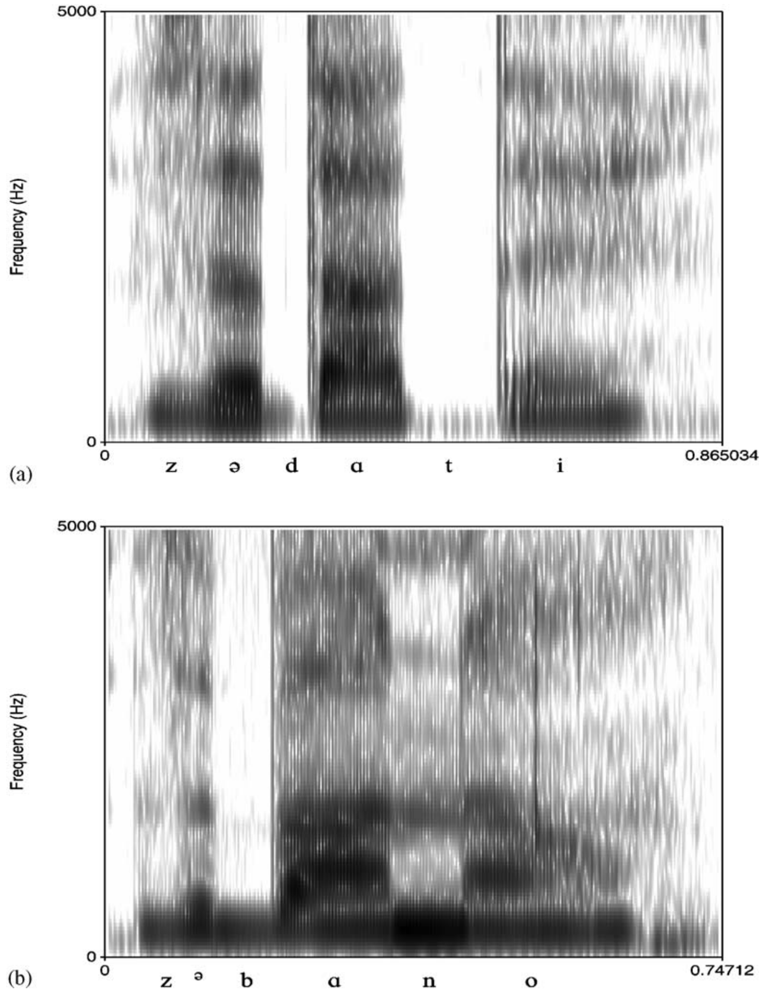

Fig. 7. Spectrograms of different manifestations of inserted vowel between voiced consonants. The top panel represents a long, intense period of voicing in inter-consonantal position of the word /zdati/ → [zədati] (subject 20). In the bottom panel, the period of voicing is shorter and weaker in the word /zbano/ → [z^{ə}bano] (subject 16).

can be produced so variably, even within a single cluster type, if it is the manifestation of a phonologically represented vowel. However, if the schwa arises from the inability to adequately coordinate the consonants, then it might be more likely for the duration and intensity of the schwa to show considerable variability, since speakers may not necessarily coordinate the consonants in the same way each time they produce a target word.

While there are many factors suggesting that the inserted vocalic material may be due to the failure of speakers to properly coordinate the consonants in the cluster, there is also another possible explanation for this variability. It may be that speakers are epenthesizing a lexical schwa, but knowing that the target is a cluster, attempt to minimize the intensity and duration of the vowel in order to produce the sequence as faithfully as possible to the target. This is similar to the concept of undershoot (Lindblom, 1963). Lindblom demonstrated for Swedish that as vowels decreased in duration, they often failed to reach their target formant frequencies. Furthermore, he argued that the directionality of the formant movement is not consistent with centralization or reduction of the vowel to schwa; rather, speakers cannot achieve the canonical target of the vowel. In the case of this study, the intentionally shorter duration of epenthesized schwas may mean that the schwas cannot attain the target formant values of a canonical lexical schwa. This leads to several predictions which will be reviewed in Section 3.

In order to provide further evidence to determine whether the period of voicing between the consonants used as experimental targets is more consistent with phonological epenthesis or "gestural mistiming", an acoustic study comparing characteristics of the transitional schwa to a phonologically intended schwa was carried out. Assuming that schwa does indeed have its own target gesture with a corresponding acoustic output (Gick, 2002; Gick & Wilson, to appear; Kondo, 1994), it can be hypothesized that a number of measures, including F1 and F2 midpoint values and duration, will differ among the two kinds of schwas, and that the mistiming and undershoot accounts lead to different predictions for schwa acoustics. An experiment designed to examine this question is presented in the next section.

3. Experiment 2

This study investigates whether the acoustic output corresponding to the repair of a non-native sequence is more consistent with epenthesis of a lexical vowel, or gestural mistiming, in which speakers do not sufficiently overlap the target consonants. In addition to the discussion in Section 2.5.2, the possibility of gestural mistiming is supported by a preliminary articulatory study using ultrasound to compare the inserted schwa (C^{ə}C) in English speakers' production of Polish /z/-initial pseudo-words to the production of the corresponding voiceless /sC/ and /səC/ sequences (Davidson, 2005; Davidson & Stone, 2004). In the study by Davidson, the production of a target like /zgomu/ as [z^{ə}gomu] was compared to the /sibilant-velar/ sequence in *scum* and /sibilant-ə-velar/ sequence in *succumb*. It was hypothesized that if the schwa on the acoustic record in [z^{ə}gomu] is a result of the failure to produce /zg/ with close coordination, then the ultrasound images of the tongue for [z^{ə}g] would be more similar to [sk] than [sək]. That is, if the schwa in [z^{ə}gomu] does not have its own gestural target, than there should be no tongue position corresponding to the vowel, as there is for *succumb*. Results showed that for 3 of 5 speakers, the tongue shape changes for [z^{ə}C] were more consistent with [sC] than with [səC], indicating that while some speakers may exhibit an epenthetic repair, the remaining speakers' productions were consistent with gestural mistiming. Since this study had only a few participants and was limited to /z/-initial sequences, the conclusions are tentative, yet they suggest that these 3 speakers were not epenthesizing a schwa gesture, but rather pulling apart the consonant gestures.

In the present study, the status of the insertion error in Experiment 1 as either a transitional or epenthetic schwa is examined by gathering acoustic information on both these repairs and on lexical schwa in English. The repaired items are compared to utterances of the same word that contain an initial [CəC] sequence (e.g., *fkada* vs. *fəkada*). The two types of schwas are compared on several variables, including F1 and F2 midpoint values, and duration of the schwa (for similar measurements of different kinds of schwas in American English, see Flemming, 2004b).

It is hypothesized that if the schwa in the repair is a result of gestural mistiming, then it should have a shorter duration than lexical schwa, since C1 and C2 will still be slightly overlapped (see Fig. 5). In addition, first and second formant (F1 and F2) midpoint values should be significantly different from those for lexical schwa since the tongue does not move toward a specified target for schwa. Specifically, it is predicted that F1 and F2 values will be lower than they are for canonical schwa. Since a transitional schwa is a brief period of open vocal tract between two constrictions (e.g., from fricatives to stops, fricatives, or nasals), the mouth will be more closed than it is for a lexical schwa, leading to lower F1 (Flemming, 2004a; Kondo, 1994; Wallace, 1994). Lower F2 may occur if the tongue root is being positioned for the first vowel after the cluster, which is an /a/ in all cases. Though the tongue tip and body will be displaced by the surrounding consonant constrictions, the tongue root is free to anticipate the upcoming vowel. If a pharyngeal constriction for /a/ begins during the cluster, it could result in a lower F2 for the transitional schwa.

If the schwa is instead a shortened, "undershot" lexical vowel, the changes in F2 relative to a full lexical schwa should be dependent on the context of the surrounding consonants. More specifically, Kondo (1994) demonstrated that F2 for schwa in /p_p/, /t_t/ and /k_k/ environments increased from /p/ to /t/ to /k/. Thus, if the speakers in the present study were undershooting the schwa target, it might be expected that F2 would move in different directions depending on the context. For example, F2 would decrease for labial or other front contexts for the inserted schwa, whereas dorsal environments would lead to an increase in F2. As for F1, some researchers have found that changes in F1 are much less frequent during undershoot (van Son & Pols, 1992), and others have found that if F1 changes at all, it increases, since ultimately the vowel space should shrink when speakers employ undershoot (e.g., Flemming, 1995/2002; Padgett & Tabain, to appear). Thus, if speakers are undershooting an epenthesized schwa, it is expected that F1 will either not be different relative to the F1 of a full lexical schwa, or will be higher.

3.1. Participants

The participants were 20 New York University undergraduates who received payment for their participation. All of them were native speakers of English. Data from two additional participants who had been exposed to Hebrew were discarded. None reported any history of speech or hearing impairments.

3.2. Materials

Half of the stimuli used in the study were the same pseudo-Czech words with /s/, /f/, /z/, and /v/-initial obstruent clusters used in Experiment 1 (e.g., *fnada, zvaba, zbano*) (CC condition). For this experiment, these stimuli were recorded by a native Slovak speaker using a Marantz PMD-670 digital solid-state recorder at a 22 kHz sampling rate. An examination of the Slovak speaker's productions showed that they contained all of the same characteristics as described for the Czech speaker in Experiment 1 (see Section 2.3). Although Slovak does not have /f + nasal/ sequences natively, the sequences contained no evidence of a transitional vocoid. The other half of the stimuli were a new set of CəC-initial stimuli (CəC condition) that were created by taking the original items and having the Slovak speaker produce them with a schwa between the initial consonants (e.g., *fənada*, *zəvaba*, *zəbano*). Again, this is not a possible Slovak pattern, but since the speaker recording the items is a proficient speaker of English and a trained phonetician, he had no trouble producing a reduced schwa vowel in that environment. For the voiceless CəC sequences, the average duration of the schwa was 48.8 ms, and the average schwa duration for the voiced sequences was 73.7 ms.

The experiment was designed with PsyScope 1.2.6 and was presented on a Macintosh G3 computer. Participants were seated in a sound-proof booth and their responses were digitally recorded at 22 kHz with the Marantz recorder using a condenser microphone.

3.3. Design and procedure

The task design is the same as that described in Experiment 1. In this experiment, the 96 CC words were interspersed with the 96 CəC words and were presented in a randomized order to each of the participants.

For this experiment, the tokens were examined by a coder blind to the purposes of the study. The CC words were coded with the same procedure described for Experiment 1. In addition, the duration and F1 and F2 at the temporal midpoint were measured in the schwa in CəC words and in the vowel (if present) between the two consonants in the CC items. All measurements were taken using Praat. The duration of the vowel was measured using the following criteria (N = nasal, O = obstruent): for /fN/, from the onset of voicing after frication to the point of abrupt lowering of intensity for the higher formants; for /fO/, from the onset to the offset of voicing; for /zN/ and /vN/, from the onset of the second formant to the point of abrupt lowering of intensity for the higher formants; and for /zO/ and /vO/ from the onset of the second formant to the cessation of F1 and F2 preceding the obstruent closure. For each of these intervals, the midpoint F1 and F2 values were obtained using linear interpolation Burg LPC, with a time step of 10 ms, window length of 25 ms, and pre-emphasis of 50 Hz. In some cases, a schwa between two voiceless consonants appeared only as a voicebar without any information for F1 and F2. Duration was noted, but no formant information was recorded for these cases.

3.4. Results

3.4.1. CC stimuli: effect of first and second consonant

The analysis of just the subset of CC stimuli demonstrates that the results from Experiment 1 have been replicated. This is confirmed with an ANOVA with the independent variables of first segment (/s/, /f/, /z/ and /v/) and second segment (obstruent vs. nasal) and average proportion of correct responses (arcsine transformed) as the dependent variable. Mean proportion correct for each first segment is presented in Fig. 8. Results show a main effect of first segment type [$F(3,57) = 108.07, p < .001$, partial $\eta^2 = .850$] and second segment type [$F(1,19) = 8.9, p < .01$, partial $\eta^2 = .319$]. Scheffé post hoc comparisons show that each of the initial segment categories were all significantly different from one another ($p < .001$). The interaction between first and second segment type was also significant [$F(3,57) = 3.32$, $p < .03$, partial $\eta^2 = .149$]. A graph of the proportion correct broken down by C1 and C2 combination is given in Fig. 9. Scheffé post hoc tests comparing each combination of C1 and C2 indicated that all categories were significantly different than one another ($p < .001$), except for /vN/ vs. /vO/.

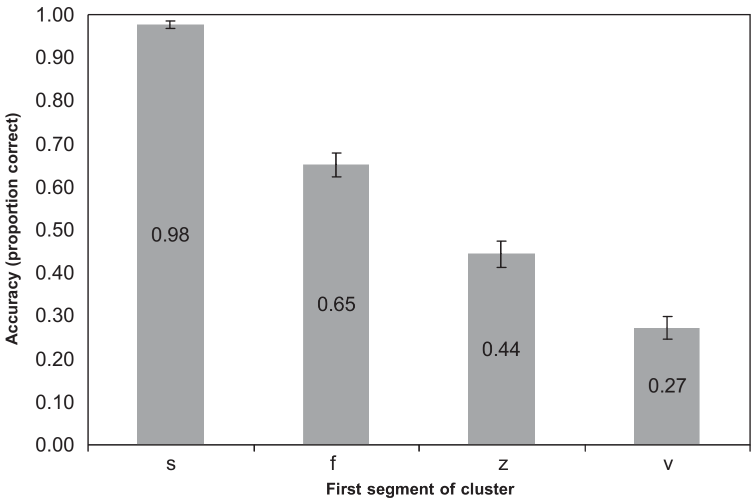

Fig. 8. Proportion correct on each cluster type plotted by first segment (Experiment 2). Error bars indicate standard error.

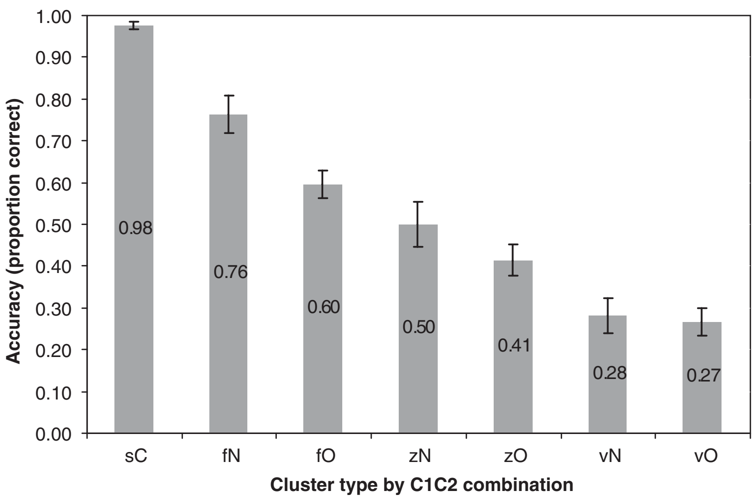

Fig. 9. Proportion correct on clusters broken down by cluster type (N = nasal, O = obstruent) (Experiment 2). Error bars indicate standard error.

The results for C1C2 cluster type differ slightly from Experiment 1 in which some CN and CO combinations were produced with equal accuracy. The differences between the two experiments may be because the variance for the individual clusters was slightly greater in the sample in Experiment 1, so all of the differences could not be detected. Ultimately, the pattern in Experiment 1 is compatible with the results in Fig. 9, and both experiments show that the combination of C1 and C2 contributes to the overall performance on the cluster.

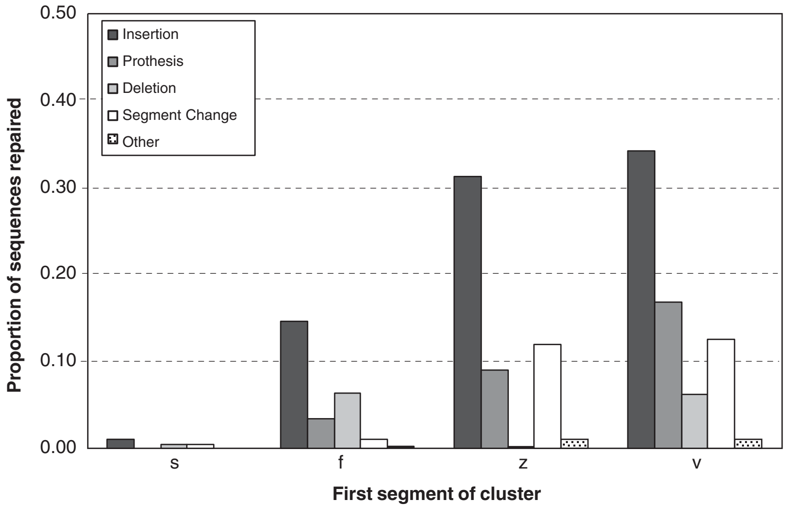

Fig. 10. Proportions of error types for each first segment type, averaged over all participants.

3.4.2. *Error types*

The distribution of error types is shown in Fig. 10. Again, insertion was the most prominent error when speakers are unable to accurately produce the target non-native cluster.

In this experiment, speakers tended to exhibit higher proportions of other repair types than they did in Experiment 1. One explanation for the change is that in this study, participants might have been trying to distinguish between their production of CC-initial words and CəC-initial ones. That is, speakers often did not accurately produce the clusters, but since there were both CC and CəC words in the stimulus set, one way to discriminate between which words were being produced was to repair CC words with something other than insertion.

The data suggest that even within a single speaker, more than one repair type is possible when speakers are faced with producing non-native sequences. For the purposes of this study, however, the repair of interest is insertion; it is beyond the scope of this study to investigate why speakers do not consistently use the same repair. The question that is addressed in the next section is the following: when speakers produce /CC/ targets as [C^{ə}C] (where superscript indicates that the schwa is a repair of the /CC/ input), are they epenthesizing a schwa or mistiming the gestures (or perhaps using both strategies)? To answer this question, the acoustic parameters of duration, F1 midpoint, and F2 midpoint are assessed with respect to the predictions presented in Section 3.

3.4.3. *Duration, F1 and F2 midpoint values*

Since participants did not make insertion errors on every CC token, only the CəC tokens corresponding to the CC words that were produced with insertion errors were compared in the analyses of duration and F1 and F2 midpoint values. That is, if a participant produced /vgadi/ and /vgalu/ as [v^{ə}gadi] and [v^{ə}galu] but produced /vgapo/ as [gapo], then only /vəgadi/ and /vəgalu/ were included in the tokens from the CəC condition that were employed in the statistical analysis. The tokens to be included were determined for each participant individually, but

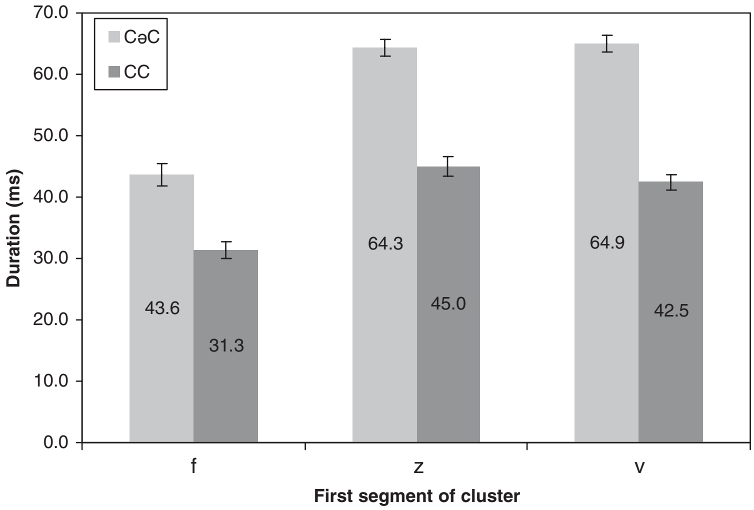

Fig. 11. Duration differences for lexical schwa (CəC condition) and transitional schwa ($C^{ə}C$ condition). Error bars indicate standard error.

responses for each of the 18 individual CC combinations making up the /fC/ stimuli (79 total tokens), /zC/ stimuli (133 tokens), and /vC/ stimuli (137 tokens) were averaged across speakers in order to conduct a within-subjects ANOVA (/s/-initial sequences were not analyzed since there were so few insertion errors on CC tokens).

Durational differences between CC and CəC tokens were analyzed with an ANOVA with the independent variables being initial segment type (/f/, /z/, /v/) and condition ($C^{ə}C$ or CəC). Duration differences for the initial segment types /f/, /z/, and /v/ in the $C^{ə}C$ vs. CəC condition are shown in Fig. 11. Results show that there are main effects of both condition [$F(1,343) = 129.33, p<.001$, partial $\eta^2 = .274$] and segment type [$F(2,343) = 37.03, p<.001$, partial $\eta^2 = .178$]. The interaction between segment type and condition was not significant [$F(2,343) = 2.43, p = .09$, partial $\eta^2 = .014$]. Post hoc comparisons show that the differences in lexical vs. inserted schwa duration for each of the initial segment types were significant (/f/: $F(1,77) = 16.15, p<.001$; /z/: $F(1,131) = 66.05$, $p<.001$; /v/: $F(1,135) = 73.03, p<.001$).

Differences in F1 midpoint values were analyzed with an ANOVA. F1 values for both CC and CəC are plotted in Fig. 12. Results indicated a main effect of both condition [$F(1,324) = 43.21, p<.001$, partial $\eta^2 = .118$] and segment type [$F(2,324) = 3.61$, $p<.03$, partial $\eta^2 = .022$]. The interaction between segment type and condition was not significant [$F<1$]. Post hoc comparisons show significant differences for each initial segment types ($/f/ : F(1,62) = 13.28$, $p<.001$; /z/: $F(1,130) = 22.77$, $p<.001$; $/v/ : F(1,132) = 13.19$, $p<.001$).

F2 midpoint values for both $C^{ə}C$ and CəC are graphed in Fig. 13. An ANOVA revealed a significant main effect of condition [$F(1,324) = 10.88, p<.001$, partial $\eta^2 = .033$] and segment type [$F(2,324) = 22.85, p<.001$,partial $\eta^2 = .124$]. The interaction between segment type and condition was not significant [$F<1$]. Post hoc comparisons indicate that there were significant differences for /f/ and /z/, but not /v/, though it is in the right direction (/f/: $F(1,62) = 3.83, p<.05$; /z/: $F(1,130) = 5.07, p<.02$; $/v/ : F(1,132) = 1.95, p = .16$)

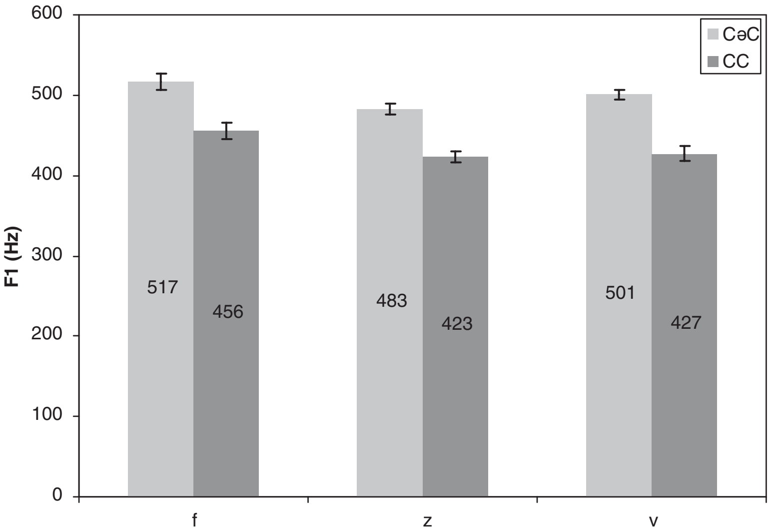

Fig. 12. F1 midpoint differences for lexical schwa (CəC condition) and transitional schwa ($C^{ə}C$ condition). Error bars indicate standard error.

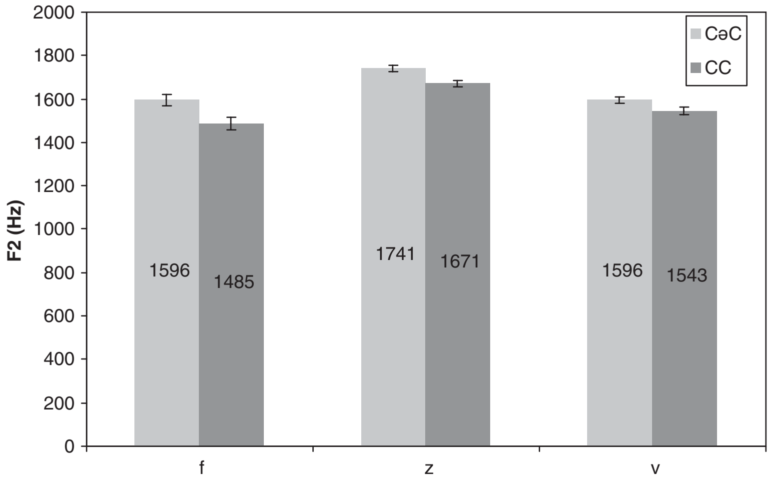

Fig. 13. F2 midpoint differences for lexical schwa (CəC condition) and transitional schwa ($C^{ə}C$ condition). Error bars indicate standard error.

3.5. Discussion

The first part of Experiment 2 demonstrates that the findings of Experiment 1 have been replicated. The results for the CC stimuli confirm that English speakers were more accurate in the production of /fC/ word-initial clusters, followed by /zC/ and then /vC/. Furthermore, the effect of C2 once again established that for each C1, speakers were more accurate on /CN/ clusters than they were on /CO/. A comparison of all clusters, shown in Fig. 9, indicates that there is a cumulative effect of the combination of C1s and C2s leading to progressively decreased accuracy from /fN/ $\succ$ /fO/ $\succ$ /zN/ $\succ$ /zO/ $\succ$ /vN/, /vO/. A comparison of the experiments on the cluster breakdown suggests that the results of Experiment 2 are generally more sensitive to the cumulative effects of C1 and C2, although Experiment 1 is not inconsistent with these findings.

The error types for CC stimuli in Experiment 2 were also similar to those for Experiment 1, although there were slightly more instances of repair types other than insertion in Experiment 2. As noted above, one possible reason for the difference is that because the stimuli in Experiment 2 contained CC and CəC words, speakers may have decreased the number of insertion repairs of CC targets in an attempt to differentiate between their productions of CC and CəC items. Despite the acoustic differences between lexical and inserted schwa as indicated by the second part of Experiment 2, it is possible that a listener hearing a transitional schwa as produced by the native English-speaking participant would reanalyze it as a lexical schwa, or a phonological segment intended by the speaker (see Hall, 2003; Kang, 2003 for more on the role of perception in loanword adaptation). Consequently, speakers may have reason to avoid such a confusion by employing a different repair, even though other repairs often sacrifice information about the original target in a way that the insertion error does not. Note however, that insertion is still the most common repair.

The analysis of the CəC stimuli confirms that the acoustic characteristics of C^{ə}C productions are qualitatively different from CəC. Regardless of whether the first consonant is /f/, /z/, or /v/, the duration of all C^{ə}C tokens is significantly shorter than CəC, and both F1 and F2 midpoint values are lower. These findings are consistent with a transitional schwa that does not have a gestural target and that is present on the acoustic record because the speakers have failed to produce the surrounding consonants with a sufficiently overlapping coordination. As pointed out in the predictions at the beginning of this section, lower F1 is consistent with a short period of unobstructed vocal tract with a more closed mouth position than a schwa target would have.[3] This configuration is expected when both consonants surrounding the transitional schwa in a C^{ə}C tokens are constricted gestures such stops or fricatives. That is, in the production of slightly overlapped consonants that are not broken up by a schwa target, the tongue tip, blade and body move from one constriction location to another, without ever having to lower enough to achieve the more open vocal tract corresponding to a vowel, even when that vowel is schwa. The lower F2 is consistent with a pharyngeal anticipation of the /a/ following the cluster. If the tongue root begins a low back constriction during production of the consonants, it would lead to a lower F2 for the transitional schwa than for the lexical schwa.

The other possible reason that the inserted schwa might differ in quality from the lexical schwa is that speakers attempt to produce the phonotactically illegal sequences as faithfully as possible to the intended sequences. Under this interpretation, speakers hear the difference between CəC and C^{ə}C and accordingly produce a maximally short schwa. In this scenario, which has been described as undershoot, the inserted schwa could still be epenthetic, but intentionally shortened.

[3]As noted by Kondo (1994) on the basis of acoustic information and Gick (2002) using X-ray data, schwa in American English does seem to have a phonological and phonetic target. Kondo's results suggest that schwa has a definable F1 target value, but that F2 is essentially determined by the surrounding consonants. She concludes that schwa is specified for height, but perhaps not for backness. Gick demonstrates that schwa has a canonical articulatory shape that is defined by some reaction of the tongue root, though not nearly as much retraction as is found for /a/.

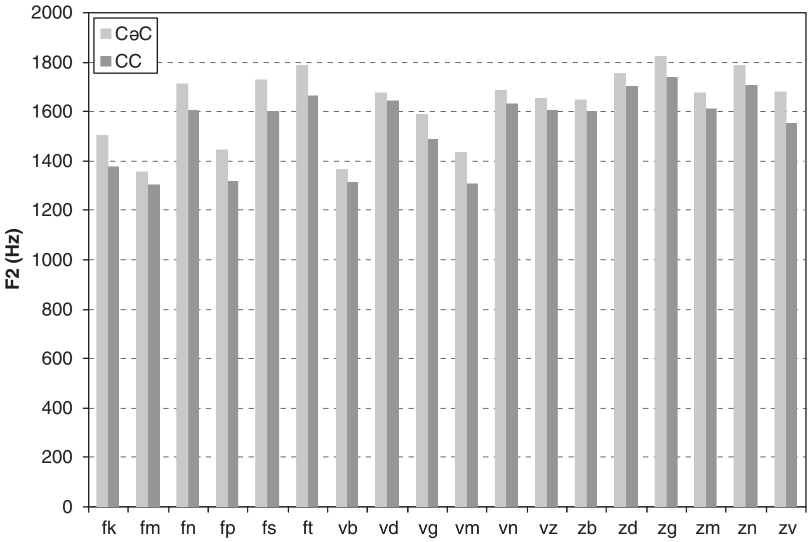

Fig. 14. F2 midpoint differences for lexical schwa (CəC condition) and transitional schwa ($C^{ə}C$ condition) for individual cluster types.

However, the analysis of F1 and F2 midpoint values is not consistent with undershoot. In Lindblom's model, acoustic undershoot is most likely to be observed when there is a large locus–target distance. In other words, undershoot only occurs when speakers must make a drastic movement from the tongue position of a consonant to the position of a vowel that is very different. Since the articulation of schwa usually accommodates to the surrounding consonants, it is unlikely that the locus–target distances for the stimuli in this study are long enough for undershoot to occur, and yet Experiment 2 contains significant differences for F1 and F2 midpoint values. Furthermore, even if there is undershoot for schwas, the direction of change for F2 for an undershot schwa should not be consistently lower regardless of the surrounding consonants, but should rather decrease in anterior contexts and increase in posterior contexts. However, as shown in the graphs for individual clusters in Fig. 14, F2 is lower for all consonantal contexts. Undershoot cannot explain this pattern, but it is consistent with the hypothesis that the tongue root is anticipating the upcoming /a/, so the transitional schwa is produced with a greater pharyngeal constriction, lowering F2.

As discussed in the predictions, if there is undershoot, then either F1 should not change or should rise. In this study, however, F1 is significantly lower for all consonant combinations, as shown in Fig. 15. Such a finding is expected in gestural mistiming, since articulators are moving from one constriction to another, with the tongue moving from the palate only long enough to allow for a short transitional schwa. In sum, the results of this study are not consistent with the theory of undershoot, but they conform to the predictions of a gestural mistiming account.

Finally, we can also verify that the inserted schwas in Experiment 2 were not artificially shortened in order to distinguish them from the lexical schwas by comparing their duration to those for the inserted schwas from Experiment 1. That is, the differences between inserted and lexical schwas could be a task effect: because speakers were given tokens with both clusters and

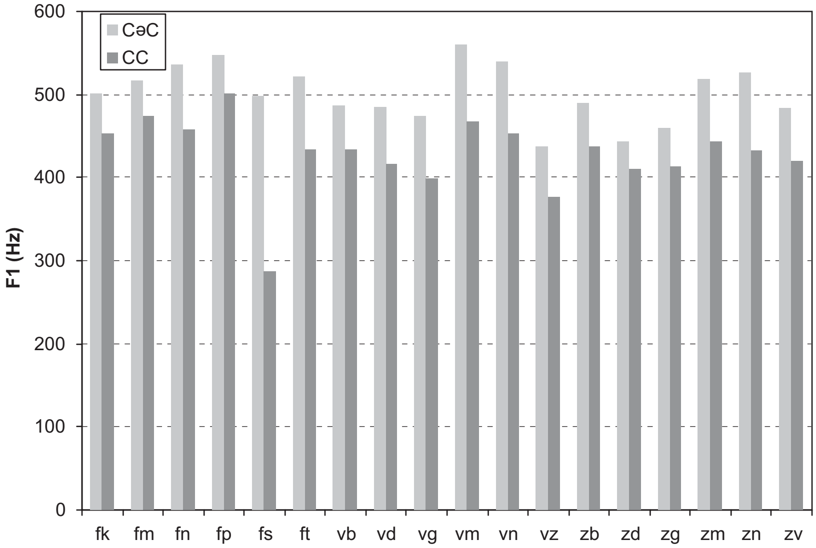

Fig. 15. F1 midpoint differences for lexical schwa (CəC condition) and transitional schwa (C^{ə}C condition) for individual cluster types.

Table 4
Lexical and inserted schwa durations from Experiment 2 compared to inserted schwa durations from Experiment 1, divided by cluster type

	Experiment 2		Experiment 1
	Lexical	Inserted	Inserted
/fC/	43.6ms (17)	31.3ms (11)	33.7ms (12)
/zC/	64.3ms (16)	45.0ms (15)	44.0ms (20)
/vC/	64.9ms (15)	42.5ms (14)	48.7ms (19)

Standard deviation is in parentheses.

lexical schwas, it is possible that they were reducing the inserted schwas in order to differentiate the two types of sequences. To test this possibility, we may take the duration of the lexical schwas in Experiment 2 as a baseline duration for pre-tonic schwa. If speakers were actually producing epenthetic schwas but attempting to distinguish them from the lexical schwas in Experiment 2, then we would expect that the inserted schwas in Experiment 1, which were not contrasted with any other kind of output for the participants, would have a duration closer to that of the Experiment 2 baseline.

The durations of inserted schwa from Experiment 1 are similar to the duration of the inserted schwa reported in Section 3.4.3. The comparison between Experiment 1 and Experiment 2 is shown in Table 4.

The results in Table 4 demonstrate that the inserted schwas from Experiment 2 were comparable in duration to the inserted schwas from Experiment 1. This suggests that speakers were not artificially shortening the inserted schwas only in the task where they are trying to distinguish them from lexical schwas. Instead, it may be hypothesized that the repairs of both experiments are due to gestural mistiming.

4. General discussion

4.1. Phonology in the production of non-native phonotactics

The results from both experiments illustrate that when producing word-initial sequences that are excluded phonotactically in English, speakers are not equally accurate on the various cluster types. The productions of such clusters by native English speakers are potentially influenced by three factors: (1) speakers' knowledge of the lexical frequency of the clusters across all positions in other words, (2) articulatory complexity, such that /fC/ sequences are somehow articulatorily simpler than /zC/ or /vC/ sequences, or (3) analogy with phonotactically permissible sequences in English, which determines the sequences that are more easily produced and acquired.

The relationship between the accuracy in production of the sequences and the frequency of such sequences as found in other parts of English lexical items demonstrates that standard lexical frequency statistics do not explain speakers' performance. Frequency was calculated over type and token, and mono- and multimorphemic lexical items in an effort to determine which particular statistics, if any, might influence the production of non-native sequences. None of these frequency counts were significantly correlated to the production results.

This finding leaves two other possible explanations for the differences in accuracy. The first is that performance is due to articulatory variables alone. Even a quick examination of English phonotactics demonstrates that a solely articulatory origin is unlikely. To lend support to a purely articulatory basis for the English speakers' performance, it would have to be the case that English speakers have no practice at all with the motor patterns necessary to produce /fC/, /zC/ and /vC/ with overlapping coordination. Yet, English allows most of the experimental clusters word-medially (*Ma[zd]a*, *hu[zb]and*, *a[ft]er*, *hal[fp]ipe*, *lo[vb]ird*, etc.) and sometimes word-finally as well. Though there have not been any articulatory studies specifically looking at the articulation and coordination of word-medial /fricative + obstruent/ sequences, there is evidence to suggest that these sequences are produced with considerable overlap and no transitional schwa. In an acoustic study of medial /stop + stop/ sequences, Henderson and Repp (1982) found that speakers only released the first stop about half the time, and when stops were released, "bursts appeared as distinct spikes of a few milliseconds duration (74)". Likewise, though Byrd (1996a) did not specifically examine medial sequences, her electropalatography results indicate that speakers substantially overlap the consonants in obstruent clusters, even across word boundaries. It is reasonable to assume that at least as much overlap, if not more, is also found word-medially. Consequently, if performance were based only on articulatory factors, speakers should be able to recruit the coordination and motor patterns of medial and final fricative-initial clusters to help them produce the experimental sequences. Yet, this does not seem to be the case.

Since standard measures of frequency and articulation cannot account for the accuracy results in the production of different non-native sequences, the alternative is that speakers are accessing phonotactic knowledge from an abstract phonological level. For example, in Section 2.5.1 it was argued that English speakers may draw on the phonological knowledge that word-initial clusters can start with voiceless fricatives, as /s/-initial clusters do, and also extend the permissibility of /f/ + sonorant clusters to /f/ + obstruent clusters. These results suggest that the influence of existing phonologically legal sequences affects production in a language-specific way, and that speakers of other languages that also do not have /fC/, /zC/ and /vC/ initial sequences may not demonstrate the same behavior that English speakers do.

While the results of this study do not reflect the influence of lexical frequency alone, they are consistent with a model that posits an abstract phonological level that extracts higher-level phonotactic information from across the lexicon (Pierrehumbert, 2001, 2003). As noted by Pierrehumbert (2003), only a small number of studies designed to distinguish between lexical and phonotactic effects have demonstrated that an abstract phonological level is explicitly necessary (e.g., Bailey & Hahn, 2001; Frisch & Zawaydeh, 2001; Vitevich & Luce, 1998, 1999). It seems that the present study also supports the need for an abstract level of phonological generalization, since speakers do not produce all of the unattested consonant clusters either with equal accuracy or with accuracy reflecting the distribution of the clusters in other positions in the lexicon. The differences in accuracy also support Pierrehumbert's claim that even abstract phonological knowledge is probabilistic, since the speakers showed a gradient increase in accuracy from /vC/ to /zC/ to /fC/ initial sequences. The nature of the mechanism used by the phonological grammar to extract information from the lexical level is still being explored, but it has been hypothesized in the present study that analogy, which has been used to describe how speakers generalize across morphologically related words (e.g., Baayen, 2003; Bybee, 2001; Skousen, 1989), may be available to speakers producing unattested phonotactic patterns.

The present study also provides evidence for the type of information that the phonological level accesses for analogical extension. Specifically, the results suggest that speakers are sensitive to a featural level of phonological encoding, since they are more accurate on the clusters that share more features with phonotactically legal sequences. For example, performance on /fC/ sequences may be relatively high not only because of the existence of /f/ + sonorant sequences, but also because they are only one place feature removed from well-attested /s/ + nasal and /s/ + obstruents sequences. On the other hand, /v/ + obstruent sequences have no similar /v/ + sonorant counterpart, nor are /z/ + obstruent sequences permitted in English. A number of studies have shown that speakers and hearers are also sensitive to featural encoding in other domains, such as speech errors (Goldrick, 2004), perception of ambiguous speech sounds (Moreton, 2002), and infant word segmentation (Jusczyk, Hohne, & Bauman, 1999). Findings like these and the present study indicate that the abstract phonological level in a probabilistic phonology must not only be able to extract phonotactic patterns from the transitional probabilities in the lexicon, but must also include featural information in the encoding of these patterns. As Goldrick (2004) noted in his study of featural effects in speech errors, either categorical units of contrast like distinctive features or the gestural representations of Articulatory Phonology could potentially underlie the featural effects seen in speech production tasks. Given the results of Experiment 2, it seems likely that gestural representations can best account for many different aspects of speech production.

4.2. Gestural mistiming, undershoot, and epenthesis

The results of both experiments indicate that the predominant repair type for non-native clusters is vowel insertion. While it is usually assumed that the presence of a schwa in the acoustic record indicates that speakers repair the phonotactically illegal sequence by epenthesizing a schwa, the findings of Experiment 2 are consistent with the hypothesis that schwas can result from gestural mistiming. Gestural mistiming was defined as the failure of a speaker to coordinate the consonants of the cluster with sufficient overlap to prevent the production of a transitional schwa. The acoustic measures of duration, F1 midpoint, and F2 midpoint demonstrate that the lexical schwa in the CəC tokens is not phonetically equivalent to the transitional schwa in $C^{ə}C$ productions. More importantly, the types of changes—shorter duration, lower F1 and F2 midpoints—are consistent with the predictions for the transitional schwa produced by gestural mistiming. The acoustic results are compatible with Davidson's (2005) ultrasound findings that [$z^{ə}C$] productions are articulatorily more similar to [sC] sequences than they are to [səC] sequences containing a lexical schwa.

The results of Experiment 2 also cast doubt on the likelihood that the insertion repair is due to a "minimal schwa" epenthesis. It was hypothesized that speakers may be epenthesizing a schwa, but one which is reduced in duration in order to make the English speakers' output as similar as possible to the intended cluster. If the epenthesized schwa is shortened, then it is expected that it will also show effects of undershoot, such as an F2 that fails to reach its intended target. However, the acoustic findings did not conform to the predictions of the undershoot hypothesis. If speakers are undershooting, the directionality of change in F2 values as compared to the lexical schwa should vary by consonantal context, and F1 should either not be affected, or should raise if it changes at all. However, the acoustic results showed that all F1 and F2 midpoints lowered, regardless of the context. Furthermore, a comparison of the inserted schwas from the two experiments showed that their durations were comparable, indicating that the short duration the of inserted schwas in Experiment 2 was not just a task effect.

In both experiments, a range of different repair types were evident, from insertion/ gestural mistiming to segment change (e.g., *zmadu* → *smadu*) to deletion (*fpame* → *pame*). This is expected for a phonological system in flux. In attempting to accurately produce the illegal sequences, speakers may try different types of repairs. Given that multiple repair types occur, it is also expected that some speakers should exhibit actual schwa epenthesis, not just gestural mistiming. It could be that the majority of some speakers' insertion repairs speakers were cases of epenthesis, while other speakers' productions were the result of gestural mistiming. If epenthesis occurs, then the duration, F1 and F2 values for those speakers should be quite similar for each cluster type in the CəC/$C^{ə}C$ comparison. An examination of individual subject data from Experiment 2 demonstrates that overall, speakers conform to the general pattern; in the few cases that a speaker's data does not follow the group pattern, it is limited to one measure for one initial segment type (e.g., for Subject 8, F2 for /z/-initial sequences is nearly identical for CəC and $C^{ə}C$.) Thus, while none of the speakers may have exclusively epenthesized a schwa instead of gestural mistiming, it is possible that individual tokens could have been repaired this way.

5. Conclusion

The results from this study have shown that in a production task, English speakers distinguish among non-native word-initial clusters—/fC/, /zC/, and /vC/—which are not attested in English. Since production accuracy does not correlate with frequency across other positions in the lexicon, these findings support models that include an abstract phonological level in the grammar. In order to produce the unattested sequences, English speakers seem to be accessing phonological information about similar sequences that are phonotactically legal in word-initial position and extending their knowledge of those sequences by analogy to the unfamiliar ones. It has been hypothesized that speakers are more accurate on the experimental clusters that share more features with existing sequences, suggesting that an abstract phonological level that extracts information from across the lexicon must encode featural information.

The combination of speakers' accuracy on phonotactically illegal sequences and the gestural timing repair raises interesting issues regarding the nature of phonotactic restrictions. One way to interpret a phonotactic restriction on adjacent segments is as a prohibition banning the close coordination of specified types of the segments. In other words, the task of the speaker producing non-native consonant clusters is two-fold: not only must speakers determine whether or not the adjacent consonants are able to form a unit, but they must also assign the proper coordination in order to accurately produce the sequence. A phonological system may allow close coordination in certain positions, such as word-medially (e.g., *frisbee* for English), while banning it from others (e.g., **Zbigniew*).

When the prohibition of a sequence is conceived of as a restriction on close coordination, the gestural mistiming repair employed in the study is not surprising. The target sequence should be produced as a closely coordinated unit, but the phonology prohibits speakers from accurately doing so. The open coordination that the speakers use under these conditions has been dubbed gestural mistiming since it reflects an inability to achieve the appropriate coordination for initial consonant clusters either in English or in the Slavic target clusters. However, open coordination does not just occur in laboratory speech. On the contrary, open transitions are the default coordination in a number of languages, including Moroccan Arabic, Sierra Popoluca, Piro, Oscan, etc. (Elson, 1956; Gafos, 2002; Hall, 2003; Matteson & Pike, 1958). Thus, the repair produced by the speakers in this study is a phonological option that emerges in English under experimental conditions and perhaps in second language acquisition.

Acknowledgments

I would like to thank Stefan Benus, Matt Goldrick, and Adamantios Gafos for their insightful comments and help with this manuscript. Reviewers Ben Munson, Jen Hay and Donca Steriade and editor Jonathan Harrington from the Journal of Phonetics also provided many suggestions that have helped me improve this paper. Many thanks to Jeris Brunette for analyzing data for the second experiment. Parts of the first experiment was presented at ICPhS 2003 (Barcelona) and LabPhon 9 (University of Illinois) and was submitted as part of the author's unpublished doctoral dissertation (Johns Hopkins University, 2003). The second experiment was presented at the Boston University Conference on Language Development. Part of this research was supported by the IGERT program in the Cognitive Science of Language at Johns Hopkins University, National Science Foundation Grant 997280.

Appendix. CC stimuli used in Experiment 1

Note: Stimuli are written in IPA. Experiment 2 included these words in the CC condition, and the same words produced with a schwa between the initial consonants for the CəC condition.

/s/-initial	sm	smava	smani	smagu	smapo
	sn	snabu	snadi	snale	snava
	sf	sfano	sfadu	sfagi	sfatu
	sp	spagi	spanu	spato	spama
	st	stamo	staka	stape	staga
	sk	skabu	skafu	skadi	skamo
/f/-initial	fm	fmatu	fmake	fmapa	fmale
	fn	fnagu	fnada	fnaʃa	fnapi
	fs	fsaga	fsake	fsapi	fsalu
	fp	fpami	fpala	fpaze	fpaku
	fk	fkada	fkabe	fkati	fkale
	ft	ftake	ftapi	ftano	ftani
/z/-initial	zm	zmafo	zmagu	zmapi	zmadu
	zn	znagi	znafe	znaʃo	znase
	zv	zvato	zvabu	zvami	zvapa
	zb	zbatu	zbasi	zbagi	zbano
	zd	zdanu	zdaba	zdapo	zdati
	zg	zgano	zgame	zgaba	zgade
/v/-initial	vm	vmala	vmape	vmabu	vmadu
	vn	vnali	vnake	vnapa	vnaze
	vz	vzaku	vzamo	vzaba	vzagi
	vb	vbagu	vbano	vbadu	vbaki
	vd	vdale	vdapi	vdaba	vdagu
	vg	vgane	vgalu	vgapo	vgadi

References

Abrahamsson, N. (2003). Development and recoverability of L2 codas. *Studies in Second Language Acquisition*, *25*(3), 313–349.

Anderson, J. (1987). The markedness differential hypothesis and syllable structure difficulty. In G. Ioup, & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge, MA: Newbury House.

Baayen, R. H. (2003). Probabilistic approaches to morphology. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics*. Cambridge: MIT Press.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods. *Journal of Memory and Language*, *44*, 568–591.

Benus, S., Smorodinsky, I., & Gafos, A. (2004). Gestural coordination and the distribution of English 'geminates'. In S. Arunachalam, & T. Scheffler (Eds.), *Proceedings of the 27th annual Penn linguistics colloquium*. Philadelphia: Penn Linguistics Club.

Boersma, P., & Weenink, D. (2004). Praat: Doing phonetics by computer. *http://www.praat.org*.
Bradley, T. (2004). Gestural timing and rhotic variation in Spanish codas. In T. L. Face (Ed.), *Laboratory approaches to Spanish phonology* (pp. 197–224). Berlin: Mouton de Gruyter.
Broselow, E. (1983). Nonobvious transfer: On predicting epenthesis error. In S. Gass, & L. Selinker (Eds.), *Language transfer in language learning*. Rowley, MA: Newbury House.
Broselow, E. (1991). *The structure of fricative-stop onsets*. Unpublished manuscript, Santa Cruz.
Broselow, E., Chen, S., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261–280.
Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research*, *7*(1), 35–59.
Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*, 219–252.
Browman, C., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston, & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press.
Browman, C., & Goldstein, L. (1992a). Articulatory phonology: An overview. *Phonetica*, *49*, 155–180.
Browman, C., & Goldstein, L. (1992b). ''Targetless'' schwa: An articulatory analysis. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody*. Cambridge: Cambridge University Press.
Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*, 209–244.
Byrd, D. (1996b). A phase window framework for articulatory timing. *Phonology*, *13*, 139–169.
Carlisle, R. (1998). The acquisition of onsets in a markedness relationship: A longitudinal study. *Studies in Second Language Acquisition*, *20*, 245–260.
Catford, J. C. (1977). *Fundamental problems in phonetics*. Bloomington: Indiana University Press.
Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, *22*, 1–26.
Coetzee, A. (2004). What it means to be a loser: Non-optimal candidates in optimality theory, unpublished manuscript.
Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and their acceptability. In . *Computational phonology: Third meeting of the ACL special interest group in computational phonology* (pp. 49–56). Somerset, NJ: Association for Computational Linguistics.
Davidson, L. (2003). The atoms of phonological representation: Gestures, coordination and perceptual features in consonant cluster phonotactics. Unpublished PhD dissertation, Johns Hopkins University.
Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics*.
Davidson, L., Jusczyk, P., & Smolensky, P. (2004). The initial and final states: Theoretical implications and experimental explorations of richness of the base. In R. Kager, W. Zonneveld, & J. Pater (Eds.), *Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press.
Davidson, L., & Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production. In M. Tsujimura, & G. Garding (Eds.), *Proceedings of the West Coast conference on formal linguistics (WCCFL) 22*. Somerville, MA: Cascadilla Press.
Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578.
Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, *16*(5/6), 491–505.
Eckman, F., & Iverson, G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research*, *9*, 234–252.
Edwards, J., Beckman, M., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects of children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research*, *47*(2), 421–436.
Elson, B. (1956). Sierra Popoluca syllable structure. *International Journal of American Linguistics*, *13*, 13–17.
Flege, J. E., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125–3134.
Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics*, *7*, 71–116.

Flemming, E. (1995/2002). *Auditory representations in phonology*. New York: Routledge.
Flemming, E. (2004a). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *The phonetic bases of markedness*. Cambridge: Cambridge University Press.
Flemming, E. (2004b). Rosa's roses: Reduced vowels in American English. *Journal of the Acoustical Society of America*, *115*(5), 2541.
Frisch, S., Large, N., & Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*, 481–496.
Frisch, S., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language*, *77*, 91–106.
Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, *20*(2), 269–337.
Gick, B. (2002). An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica*, *59*(1), 38–48.
Gick, B., & Wilson, I. (to appear). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. In *Papers in laboratory phonology VIII*. Cambridge: Cambridge University Press.
Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, *51*(4), 586–603.
Hall, N. (2003). *Gestures and segments*: *Vowel intrusion as overlap*. Unpublished PhD dissertation, University of Massachusetts, Amherst, MA.
Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (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*. Oxford: Oxford University Press.
Hancin-Bhatt, B., & Bhatt, R. (1998). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition*, *19*, 331–378.
Hansen, J. (2004). Developmental sequences in the acquisition of English L2 syllable codas. *Studies in Second Language Acquisition*, *26*, 85–124.
Haunz, C. (2003). Grammatical and non-grammatical factors in loanword adaptation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences*. Barcelona, Spain: Universitat Autónoma de Barcelona.
Hay, J., Pierrehumbert, J., & Beckman, M. (2001). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI*. Cambridge: Cambridge University Press.
Henderson, J., & Repp, B. (1982). Is a stop consonant released when followed by another stop consonant? *Phonetica*, *39*, 71–82.
Jannedy, S. (1994). Rate effects on German unstressed syllables. *OSU Working Papers in Linguistics*, *44*, 105–124.
Jusczyk, P., Hohne, E., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*, 1465–1476.
Kabak, B. (2003). The perceptual processing of second language consonant clusters, unpublished dissertation, University of Delaware.
Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology*, *20*, 219–273.
Kondo, Y. (1994). Targetless schwa: is that how we get the impression of stress-timing in English? *Proceedings of the Edinburgh linguistics department conference '94* (pp. 63–76).
Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary. *Cognition*, *50*, 329–369.
Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, *35*(11), 1773–1781.
Magen, H. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, *20*, 381–400.
Major, R. (1987). English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics*, *15*, 197–202.
Massaro, D., & Cohen, M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, *34*(3), 338–348.
Matteson, E., & Pike, K. (1958). Non-phonemic transition vocoids in Piro (Arawak). *Miscellanea Phonetica*, *3*, 22–30.
McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In M. Gonzales (Ed.), *Proceedings of the 24th North East linguistics society*. Somerville: Cascadilla Press.

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, *84*, 55–71.
Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language and Hearing Research*, *44*(4), 778–792.
Padgett, J., & Tabain, M. (to appear). Adaptive dispersion theory and phonological vowel reduction in Russian, *Phonetica*.
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins.
Pierrehumbert, J. (2003). Probabilistic phonology: discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics*. Cambridge: MIT Press.
Pitt, M. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, *60*(6), 941–951.
Pitt, M., & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, *39*, 347–370.
Price, P. J. (1980). Sonority and syllabicity: Acoustic correlates of perception. *Phonetica*, *37*, 327–343.
Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer.
Smorodinsky, I. (2002). *Schwas with and without active control.* Unpublished PhD dissertation, Yale University, New Haven.
Storkel, H. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language and Hearing Research*, *44*, 1321–1337.
Storkel, H., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics and Phonetics*, *14*, 407–425.
Tajima, K., Port, R., & Dalby, J. (1997). Effects of termoral correction on intelligibility of foreign accented English. *Journal of Phonetics*, *25*, 1–24.
Tarone, E. (1987). Some influences on the syllable structure of interlanguage phonology. In G. Ioup, & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge: Newbury House Publishers.
Ussishkin, A., & Wedel, A. (2003). Gestural motor programs and the nature of phonotactic restrictions: Evidence from loanword phonology. In: G. Garding, M. Tsujimura (Eds.), *Proceedings of the West Coast conference on formal linguistics*, Vol. 22, UC, San Diego (pp. 505–518). Somerville, MA: Cascadilla Press.
van Son, R., & Pols, L. (1992). Formant movements in Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America*, *92*, 121–127.
Vitevich, M., & Luce, P. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, *9*, 325–329.
Vitevich, M., & Luce, P. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*, 374–408.
Vitevich, M., & Luce, P. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. *Journal of Memory and Language*, *52*(2), 193–204.
Vitevich, M., Luce, P., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*, 47–62.
Wallace, K. (1994). *An acoustic study of American English Schwa in multiple speaking modes.* Unpublished PhD dissertation, New York University.
Weinberger, S. (1994). Functional and phonetic constraints on second language phonology. In M. Yavas (Ed.), *First and second language phonology* (pp. 283–302). San Diego: Singular Publishing Group.
Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary. Version 2. *Behavioural Research Methods, Instruments and Computers*, *20*, 6–11.
Wonderly, W. (1951). Zoque II: Phonemes and morphophonemes. *International Journal of American Linguistics*, *17*, 105–123.
Zsiga, E. (2003). Articulatory timing in a second language: Evidence from Russian and English. *Studies in Second Language Acquisition*, *25*, 399–432.