Epenthesis Versus Gestural Mistiming in Consonant Cluster Production: An Ultrasound Study

Lisa Davidson & Maureen Stone
Johns Hopkins University & University of Maryland at Baltimore

1. Introduction

In studies of production and second language acquisition, it is typically assumed that when speakers produce a vowel between the consonants in a sequence that is phonotactically illegal in the native langauge, it is a result of the phonological epenthesis of a vowel (e.g. Tarone 1987, Broselow and Finer 1991, Hancin-Bhatt and Bhatt 1998, Davidson, Jusczyk and Smolensky 2003). For example, Tarone (1987) reported that Korean speakers learning English repaired [stop+liquid] clusters by epenthesizing a schwa between the two consonants, e.g. *class* → [kəlæs].

However, the assumption that schwa results from phonological vowel epenthesis has been indirectly challenged by some of the research in the Articulatory Phonology framework. It has been proposed that schwas in English, even ones that are generally accepted to be present underlyingly (as in *p*[ə]rade or *t*[ə]*morrow*), do not necessarily need to have their own gesture associated with them, and can be derived acoustically from variations in the coordination and distance between the flanking consonants (Price 1980, Browman and Goldstein 1990, 1992a, b, Smorodinsky 2002).

After examining x-ray tracings of tongue movement in nonsense forms like [pipəpapə], Browman and Goldstein (1992b) ultimately conceded that the tongue position for English schwa was not simply a smooth interpolation between the flanking consonants and vowels, and that it likely does have its own gestural target. However, though English schwas may not be "targetless", other languages, such as Sierra Popoluca, Piro, and

* We would like to thank the members of the Vocal Tract Visualization Lab at UMB, JHU Cognitive Science, and the audience at WCCFL for comments on this work. We especially thank Melissa Epstein for her assistance with running participants and Vijay Parthasarathy, Min Li and Chandra Kambhamettu for technical assistance. This work was funded by NIH grant R01 DC01758.

 WCCFL 22 Proceedings, ed. M. Tsujimura and G. Garding, pp. XX-XX. Somerville, MA: Cascadilla Press.

Moroccan Arabic have been analyzed as having schwas that arise not from epenthesis, but rather from consonantal gestures in clusters that are not sufficiently overlapped (Elson 1956, Matteson and Pike 1958, Gafos 2002). In other words, a salient release is produced with a vocalic aspect. In Piro, for example, an excrescent schwa, which Matteson and Pike (1958) argue to be non-phonemic, can optionally occur between the consonants. This is illustrated in (1).

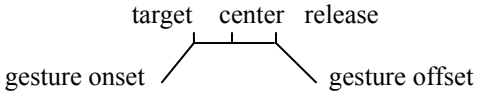
(1) /tkatʃi/ → [tᵊkatʃi]
'sun'

Excrescent schwa as characterized both by experimental data and the analyses of languages such as Piro or Moroccan Arabic presents an alternative to the assumption that the schwa in the production of non-native sequences is the epenthesis of a vowel. While it is true that the phonological process of epenthesis may repair a phonotactically illegal sequence, it is also conceivable that in a production task (by English speakers), the repair actually concerns the coordination of gestures. The cross-linguistic data suggests that for many languages, excrescent schwa arising from non-overlapping coordination is a robust phonological option. Whereas lexical schwa likely has an underlying target, a production task in which speakers are given consonant sequences for which they do not have coordination patterns may lead to a different kind of schwa. It is plausible that speakers attempting to correctly produce these sequences may nevertheless fail, but do so by using non-overlapping coordination to repair the illegal sequences. This question is addressed with an ultrasound experiment in Section 3.

2. Coordination of initial consonant clusters

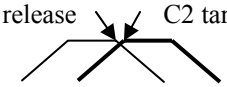
Within Articulatory Phonology, it has been proposed that differences between languages like English, which have a close transition between the members of a consonant cluster (Catford 1988), and languages like Moroccan Arabic which have transitional vocoids, are dependent on language specific coordination relations among adjacent gestures (Browman and Goldstein 1992a). Gafos (2002) proposes a framework in which to account for these cross-linguistic differences in gestural coordination. In Gafos's proposal, gestures have temporal landmarks, and the coordination relationships of adjacent gestures are defined in terms of landmarks. The temporal landmarks of a gesture are shown in (2).

(2)



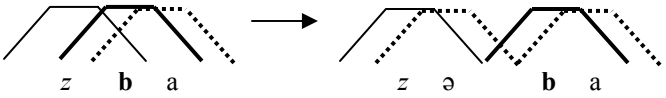
In English, close transition between consonants in a cluster occurs when the formation of the stricture of the second consonant (C2) is simultaneous with or precedes the release of the stricture of the first consonant (C1). Consequently, C1 is not audibly released. In terms of temporal landmarks, the release of C1 is coordinated with the target of C2, as shown schematically in (3).

(3) C1 release C2 target



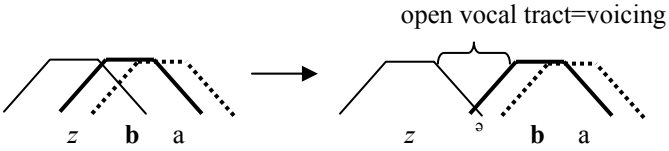
If speakers faced with phonotactically illegal consonant clusters repair them with phonological epenthesis, then it must be assumed that they are inserting a gesture corresponding to the schwa into the gestural score. A schematic for a hypothetical target word and repair are given in (4).

(4)



Another possibility, however, is that speakers are "mistiming" the consonantal gestures in the pronunciation of /zba/. Because speakers do not have experience with these sequences, it is plausible that they are unable to apply canonical English initial cluster coordination when attempting to produce them. It is hypothesized that experimental participants (and perhaps L2 learners) could eschew phonological epenthesis when they are actively attempting to accurately produce non-native sequences, but nevertheless fail to achieve accurate consonantal coordination. One possibility is that speakers are unable to overlap consonants altogether when the consonants in the sequence are not a legal cluster in the native language. This is assumed to be the standard coordination for Piro or Moroccan Arabic, which is perceived acoustically as having a schwa between the consonants. This type of configuration, which will be called *gestural mistiming*, is shown in (5).

(5)



In order to determine whether speakers are epenthesizing a vowel or mistiming gestures, ultrasound imaging of the tongue during speech can be used (e.g. Stone 1991, 1995, Iskarous 1998, Gick and Wilson to appear).

Ultrasound is an appealing technology for the study of speech because of good temporal (30 frames/sec) and spatial (<1mm) resolution. Furthermore, ultrasound is a non-invasive method, and it does not expose the subject to radiation. Ultrasound images are collected in real-time, showing tongue surface motion during speech.

3. Experiment

To distinguish between epenthesis and gestural mistiming, the articulations of non-native clusters with inserted schwa can be contrasted with two types of native English articulations: (a) legal word-initial clusters and (b) the corresponding sequences with a schwa. Assuming that oral gestures are not affected by laryngeal specifications, sequences differing only in voicing can be compared. For example, an English speaker may produce the pseudo-Polish word *zgama* as [zəgama]. The sequence of ultrasound frames corresponding to that speaker's production of [zəg] can be compared to the same speaker's [sk] in *scum* and [sək] in *succumb*. If speakers are repairing phonotactically illegal [zC] clusters by epenthesizing a schwa, the sequence of tongue shape changes in [zəC] should be similar to the tongue shapes for [səC]. However, if the speaker is mistiming the gestures, leading to an excrescent schwa, the tongue shape changes in [zəC] should be more similar to those produced for [sC]. More specifically, for both [səC] and [zəC], if schwa is a result of epenthesis, the tongue body may lower and retract after the coronal fricative in order to reach the schwa target and then move toward the position necessary for the following consonant (cf. Gick and Wilson to appear). If the gestures in [zəC] are pulled apart but there is no schwa target, the tongue body should smoothly move from the [z] to C, as for [sC].

3.1. Participants

The participants were 5 University of Maryland graduate students. All students were native speakers of American English; one was also a speaker of Korean. No students had been exposed to Slavic languages. None reported any history of speech or hearing impairments. All participants were paid for their time.

3.2. Materials

The target stimuli were three triads of /sC_i-/, /səC_i-/, and /zC_i-/ initial words. The /zC/-initial words were possible but non-words in Polish, so that all target stimuli could be recorded by a bilingual English-Polish speaker.

The second consonant of each member of a triad was matched for place, manner, and continuancy. To the extent possible, an effort was also made to match all members of the triad on the vowel immediately following the second consonant to minimize coarticulatory effects of the vowel on the production of the preceding consonant. This was not always possible however, since Polish vowels are only a subset of English vowels. For each triad, two possible pseudo-Polish words were constructed in order to improve the likelihood of capturing usable ultrasound images. The target words used in the experiment are shown in Table 1.

Triad	English /səC/	English /sC/	Pseudo-Polish /zC/
labial: /səp/-/sp/-/zb/	superfluous	spurt	[zbura], [zbertu]
coronal: /sət/-/st/-/zd/	satirical	steer	[zdiri], [zderu]
velar: /sək/-/sk/-/zg/	succumb	scum	[zgama], [zgomu]

Table 1. Target stimuli used in experiment.

In addition to the target words, 24 more legal words and 8 more non-words were also presented to the participants, for a total of 44 words. The stimuli were recorded by a bilingual English-Polish speaker using the Kay Elemetrics CSL at a 44.1-kHz sampling rate.

3.3. Design and Data Collection

An ultrasound machine (Acoustic Imaging, Inc., Phoenix, AZ, Model AI5200S) was used to collect midsagittal images of the tongue from the 5 speakers during the production of the target triads. A 2.0-4.0 MHz multi-frequency convex-curved linear array transducer that produced 30 wedge-shaped scans per second was used. Focal depth was 10cm. To ensure that the speaker's tongue would not shift during data collection, the speaker's head was stabilized by a specially designed head and transducer support (HATS) system (for details, see Stone and Davis 1995).

In ultrasound imaging, piezoelectric crystals in the transducer emit a beam of ultra high-frequency sound that is directed through the lingual soft-tissue. A curvilinear array of 96 crystals in the transducer fire sequentially, and the sound waves travel until they research the tongue-air boundary on the superior surface of the tongue. They reflect off the boundary, returning to the same transducer crystals, and are then processed by the computer which reconstructs a 90° wedge-shaped image of the 2-mm thick mid-sagittal slice of the tongue. In the reconstructed image, the upper surface of the tongue appears as a bright white line on a black background, as shown in Figure 1.

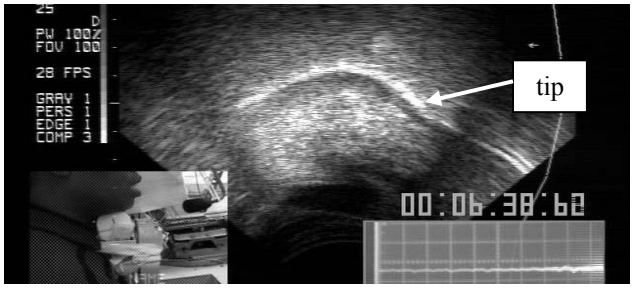


Figure 1. Sample ultrasound image of a tongue surface

Participants were seated in the HATS system, which was adjusted to fit the speaker's head comfortably. The transducer was placed under the speaker's chin and adjusted until the crispest image of the tongue was obtained. The target stimuli and filler words were presented to the speaker using PsyScope 1.2.6 on a Macintosh G3 laptop. The speakers were first given instructions informing them that they would be seeing a series of both words and non-words on the screen that would also be presented aurally. The words appeared on the screen in English-like orthography while a sample of each word, as recorded by the bilingual English-Polish speaker, was simultaneously played on an external speaker. Participants were asked to repeat each word seven times, and then wait for the experimenter's signal to move on. The whole recording procedure lasted between 15-20 minutes depending on small variations in speech rate. Both the visual ultrasound image and the synchronized acoustic signal were captured.

3.4. Methods[1]

For each token, the ultrasound frames of interest were chosen by matching the acoustic record to the ultrasound images to determine the time and duration of each /səC/, /sC/, and /zC/ sequence produced by the speaker. The middle 5 of the 7 repetitions of each sequence produced by the speaker were measured.

Visualization. Tongue shapes are measured using EdgeTrak, an automatic system for the extraction and tracking of tongue contours (Akgul, Kambhamettu and Stone 1999, Li, Kambhamettu and Stone 2002). A few points on the tongue image are chosen, and then EdgeTrak uses an active contour model to determine the location of the tongue edge in the image.

1. For an extensive discussion of the technical aspects of processing ultrasound data and more details of this study, see Davidson (2003).

After the edge of the first frame in a sequence is tracked and optimized, the algorithm is propagated to all of the tongue contours in the sequence.

Once the tongue contours are tracked, they can be displayed as a series of x, y, t surfaces using the program CAVITE (Contour Analysis and VIsualization TEchnique: V. Parthasarathy, M. Stone, J. Prince, M. Li, C. Kambhamettu, 2002-03). In order to compare repetitions of the same utterances or tongue contours that are matched for experimental variables, it must be ensured the data collection process does not introduce too much error. CAVITE is designed to minimize a number of shortcomings that may arise in extracting the tongue contours, including (a) small differences in speaking rate or mismatches in the first frame in a sequence across repetitions, (b) small spatial shifts caused by head movement, and (c) differences in tongue lengths over the course of the utterance. Using CAVITE, the multiple repetitions of a single utterance can be averaged, which is useful as a tool for visualizing the data. Once an averaged surface of each /səC/, /sC/, and /zC/ sequence is calculated, it can be displayed as a spatio-temporal XY-T surface that illustrates how the tongue changes shape over time (see (7)).

Statistical Measures. The differences between the tongue shapes for the native and non-native sequences can be quantified on a frame-by-frame basis using *L_2 norms*. The *L_2 norm* is an error metric that measures the differences between the tongue shapes for two frames based on the subtraction of the vector representing one tongue shape from the vector representing the other. The *L_2 norms* are calculated by the equation in (6) (*Y* refers to tongue height, where tongue length has already been equalized for each frame being compared). The smaller the value for the *L_2 norm*, the more similar two tongue shapes are.

$$(6) \qquad L_2 norm = \sqrt{\sum \left(Y_1 - Y_2\right)^2}$$

The sign test is used to determine whether speakers' productions of /zC/ are statistically more similar to [səC] or [sC]. To create the input for the sign test, an *L_2 norm* is calculated for each frame between every possible combination of the five repetitions for the [səC]-/zC/ comparison. The same is done for every possible combination of repetitions for the [sC]-/zC/ comparison. This generates 25 *L_2 norms* for both the [səC]-/zC/ comparison and the [sC]-/zC/ comparison for each frame, which is the input to the matched pairs sign test (see Figure 2)

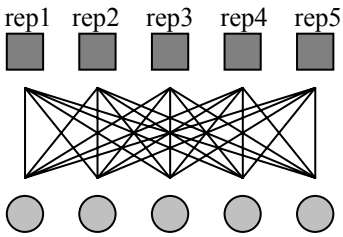


Figure 2. 25 L_2 *norms* (black lines) compare all combinations of the 5 repetitions of Frame i for [səC] (boxes) and /zC/ (circles). L_2 *norms* are calculated for each frame i (i=1-5) and matched to the corresponding L_2 *norms* for the [sC]-/zC/ comparison in the sign test calculations.

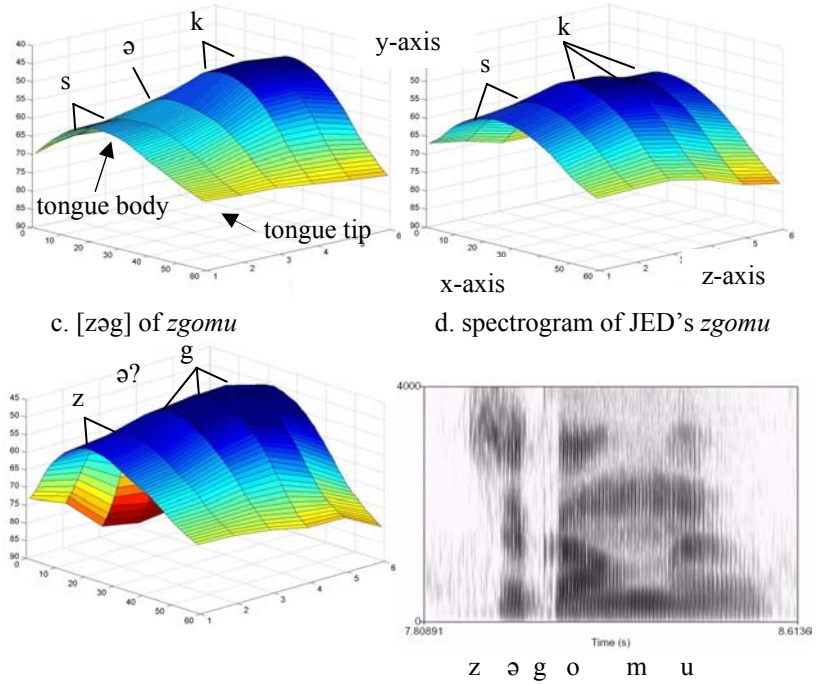
If speakers' tongue shapes for [zəC] are reliably more like [sC], then the L_2 *norms* for comparisons of the individual repetitions should be smaller for this pair significantly more often, regardless of which repetitions are compared. The sign test is a conservative test that does not assume that the data is normally distributed. For each stimulus, L_2 *norms* are calculated for first 5 frames, since it is hypothesized the major differences and similarities of /zC/ to [səC] or [sC] will be found in these frames.

For 25 comparisons, the sign test criterion values for significance are ≥18 and ≤7. That is, if the L_2 *norms* for individual repetitions are smaller for [sC]-/zC/ than [səC]-/zC/, for at least 18/25 comparisons, then it can be said that a speaker's tongue shapes for /zC/ are reliably more similar to [sC]. On the other hand, if at most 7/25 L_2 *norms* are smaller for the [sC]-/zC/ comparison, then a speaker's tongue shapes for /zC/ are significantly more similar to [səC]. A two-tailed test with an alpha value of .021 is used.

3.5. Results

To exemplify, the spatio-temporal surfaces are shown for speaker JED's velar triad (*succumb, scum, zgomu*) in (7)a-c. Note that these figures do not represent the surface of an entire tongue, but rather 6 connected frames. The x-axis represents tongue length, the y-axis is tongue height, and the z-axis is the number of frames. The units of the x- and y-axes are in millimeters. The shading reflects the height of the tongue curve: dark (blue or gray) is a high position, and lighter (blue or gray) is a lower position. The spectrogram for JED's production of *zgomu* as [zəgomu] is shown in (7)d.

(7)a. [sək] of *succumb* b. [sk] of scum



Impressionistically, a number of differences can be seen among the images in (7)a-c. First, the [s] of [sək] is characterized by a tongue body position that is lower than that of the other two tokens. This is indicated in the images by a lighter coloring (gray or blue) which corresponds to a lower position in [sək] and a darker coloring (gray or blue) which indicates a higher position for [sk] and [zəg]. More generally, the first three frames are flatter and less vaulted than those frames for [sk] or [zəg].

Note that if the schwa target in the word *succumb* corresponds to its own gesture, then it might be expected that the tongue blade and body would be slightly higher for the production of the [s], lower and perhaps retracted for the production of the [ə] (Gick 2002, Gick and Wilson to appear), and then raised considerably in the dorsal region for the production of the [k]. However, this does not appear to be the case for *succumb*. Evidence for a schwa target comes from the starting position of the [s], which coarticulates with the immediately following gesture. In *scum*, the [s] coarticulates with the [k], which has a very high tongue body position necessary for creating a velar closure. This causes the [s] to start in a

relatively high position also. The production of *scum* can be contrasted with *succumb*, in which the [s] has a considerably lower starting position because it is coarticulating with a [ə], which has a tongue body target position that is lower in the mouth. Despite the fact that the acoustic record for JED's production of /zg/ contains a schwa, the tongue shape changes for [zəg] appear more similar to those for [sk] than [sək]. Like the [s] in *scum*, the [z] appears to be coarticulating with the [g], which has a high tongue body target, rather than with a schwa gesture which would presumably force the [z] to have a lower starting position for the tongue body.[2]

Each of the five speakers produced the coronal, labial, and velar triads. The final data set including all participants' utterances contained a total of 5 repetitions each of 15 productions of the non-native word with a /zC/-initial sequence. Of these 15 productions, 11 of them contained a schwa in the acoustic record.[3] In the other cases, the non-native targets were either devoiced or produced correctly. Each speakers' production of stimuli with a /zC/ initial cluster is summarized in Table 1.[4]

JED	HJC	PDD	KAH	ELR
[zderu]	[zədiri] (71)	[zderu]	[zədiri] (62)	[steru]
[zəbura] (24)	[zəbertu] (34)	[zəbura] (37)	[zəbertu] (31)	[zəbura] (54)
[zəgomu] (35)	[zəgomu] (47)	[zəgomu] (34)	[zəgomu] (41)	[skama]

Table 1. Speakers productions of /zC/ initial clusters. The schwa duration in milliseconds, averaged over the 5 repetitions, is in parentheses.

The mean L_2 *norm* of the 25 values for each of the comparisons is shown in Table 2.[5] The sequences under the speakers' initials (column 1) indicate the speaker's pronunciation for the /zC/ targets. The two columns of numbers for each triad are the mean L_2 *norms* for each frame. The

2. Note that /z/'s tongue position is not inherently high; for example, images for *zealot* show that [z], which coarticulates with [ɛ], starts relatively low in the mouth.

3. Whenever a speaker produced a /zC/ target with a schwa, he or she was consistent throughout all of the repetitions.

4. Although each speaker produced 2 tokens for each of the /zC/ targets, only the stimulus with the best image for each speaker was measured. This is why different target words for the /zC/ stimuli are given in Table 1.

5. In ELR's production of *zderu,* PDD's production of *zbura*, and JED's production of *zgomu*, only 4 repetitions of each word were included due to measurement errors. Since this gives 20 comparisons to be submitted to the sign test, the criterion values were 5 and 15 just for these three triads.

smaller number indicates a closer pattern. Differences between the comparisons of less than 1 were considered within measurement error. The *L_2 norm* values for significantly smaller differences are shaded gray (i.e. the sign test value is ≤7 when the /səC/-/zC/ comparison is significantly smaller, and ≥18 when the /sC/-/zC/ comparison is significantly smaller).

		Cor: /sət/-/st/-/zd/		Lab: /səp/-/sp/-/zb/		Vel: /sək/-/sk/-/zg/	
	Fr	/sət/-/zd/	/st/-/zd/	/səp/-/zb/	/sp/-/zb/	/sək/-/zg/	/sk/-/zg/
ELR [st] [zəb] [sk]	1	10.4	12.0	28.8	14.5	30.7	14.4
	2	10.8	15.4	35.0	16.4	41.4	24.3
	3	11.2	19.9	49.8	17.7	37.8	26.7
	4	13.2	25.6	59.1	20.8	27.3	24.0
	5	17.9	31.3	59.3	25.0	20.8	21.7
JED [zd] [zəb] [zəg]	1	15.1	11.2	15.9	11.3	28.0	15.8
	2	11.7	13.3	17.8	12.6	35.3	17.6
	3	9.4	20.1	25.8	17.5	36.1	18.7
	4	11.8	28.6	34.7	24.7	17.4	22.9
	5	18.9	32.3	42.8	36.5	20.4	32.7
PDD [zd] [zəb] [zəg]	1	18.8	14.6	17.9	12.8	13.5	10.4
	2	18.2	10.4	22.8	13.2	22.8	15.4
	3	11.6	11.5	26.3	16.3	28.2	16.1
	4	15.1	15.9	31.5	22.0	23.2	16.6
	5	22.5	18.5	33.6	20.6	15.9	16.1
KAH [zəd] [zəb] [zəg]	1	13.2	13.8	13.8	17.6	16.6	20.1
	2	12.7	14.7	14.6	16.7	23.0	25.0
	3	11.7	14.3	15.3	14.0	32.1	29.0
	4	10.5	13.4	13.8	12.3	26.3	22.7
	5	11.2	11.8	15.9	11.8	18.6	20.3
HJC [zəd] [zəb] [zəg]	1	13.9	14.2	9.2	15.4	17.5	18.1
	2	13.3	15.1	9.6	14.9	18.8	18.0
	3	12.2	14.4	10.9	15.6	22.1	23.9
	4	11.7	13.3	11.9	17.0	26.1	26.5
	5	13.5	16.1	17.3	19.2	26.1	23.1
Significant frames		**8**	3	5	**16**	2	**13**

Table 2. Mean *L_2 norm* results for /zC/ comparisons to [səC] (odd columns) and [sC] (even columns) for each of the 5 measured frames in the sequence. Gray shading identifies a significantly smaller mean, i.e., which legal sequence is more similar to the /zC/.

The tally of significant differences (last row) shows that the coronal sequences behaved differently from the velars and labials. In the velar and

labial data, /zC/ tongue shapes were generally more similar to the [sC] sequences than [səC] sequences, despite the fact that in all cases (except ELR's velar triad), there was an acoustic schwa (column 1). For the coronals, on the other hand, there were fewer significant similarities, and they mostly indicated that the tongue shapes were similar to the [səC] shapes even though 3 subjects did not have an acoustic schwa. There was a subject effect as well. JED, ELR and PDD strongly reflected the majority pattern and accounted for much of the significant data. KAH reflected this pattern more weakly with significant comparisons mostly in the final frames. HJC produced a schwa in all contexts; her tongue shapes were more similar to [səC] for the labial case and did not show statistically greater similarity to either pattern for the velar and coronal data. With these data, 4 of the 5 subjects have tongue shapes consistent with patterns for coronals differing from velars and labials, and for non-coronals, tongue shapes do not reflect the acoustic schwa.

3.6. General Discussion

It was hypothesized in Section 3 that greater similarity of [zəC] to [sC] tongue shapes would arise if the output of the phonology for /zC/ word-initial clusters does not actually include a schwa gesture with its own target. Instead, for at least three of the speakers, the schwa present on the acoustic record follows from the hypothesis that speakers are pulling apart the /z/ and subsequent consonant to prevent their overlap, since they do not have experience with the coordination necessary for the appropriate production of a /zC/ word-initial cluster. If the vocal tract between the constrictions of the two consonants is sufficiently open, a vowel will be perceived. If a schwa target were actually present in the production of [zəC], it would have direct consequences for the production of the preceding consonant. Since the tongue shape and position of /s/ and /z/ seem highly dependent on the immediately following gesture, whether that gesture is /ə/ or a consonant has a considerable effect on the shape of the initial fricative.

The remaining two speakers do not appear to be using a uniform approach to produce the /zC/ targets. It is possible that these speakers are using multiple strategies to produce the phonotactically illegal targets. They could be using a combination of epenthesis and gestural mistiming, or they could be using an entirely different strategy that has not yet been considered. Alternatively, it is possible that there is not enough data for these speakers, and that more repetitions would provide more robust results.

The unexpected findings regarding the production of the coronal triad are likely related to the fact that the consonants in that triad are homorganic. Because the articulation of /s/ and /t/ is very similar, with the essential difference between them being a tightening of the constriction in the

alveolar region from /s/ to /t/ (or /z/ to /d/) (Catford 1988), then there is no chance for the vocal tract to be open between the two consonants unless the speaker moves the tongue away from the palate. Even if the consonants in the cluster are coordinated such that they do not overlap, the motion of the tongue from /s/ to /t/ or /z/ to /d/ will not result in an excrescent vowel unless it is pulled away from the alveolar ridge (cf. a similar discussion in Gafos 2002). For example, although the acoustic output for JED appears to be [zd], it could be that this results from a configuration which is non-overlapping. If JED does not fully pull the tongue away from the palate, then [zd] may be heard, but if the tongue wavers from the palate at all, then the production may seem more similar to [sət] articulatorily. This is consistent with the sign test results. The issues unique to the production of the coronal sequence may shed light on why there are differences in behavior on these sequences versus labial and velar.

4. Conclusions

It has previously been assumed that speakers typically repair phonotactically illegal sequences with epenthesis of a phonological vowel. However, following claims in the Articulatory Phonology literature, this study demonstrates that at least some speakers' data supports the hypothesis that speakers do not necessarily use phonological epenthesis to repair illegal sequences, but rather fail to employ the appropriate gestural coordination for English initial consonant clusters. Using ultrasound imaging, it has been shown that speakers' tongue motion during their production of /zC/ sequences is not consistent with movement toward a schwa target. Instead, this study provides evidence that some speakers are likely to be pulling apart the consonant gestures, which gives rise to a brief open vocal tract and an excrescent schwa on the acoustic record.

References

Akgul, Yusuf Sinan, Chandra Kambhamettu and Maureen Stone. 1999. Automatic extraction and tracking of the tongue contours. *IEEE Transactions on Medical Imaging* 18 (10), 1035-1045.

Broselow, Ellen and Daniel Finer. 1991. Parameter Setting in Second Language Phonology and Syntax. *Second Language Research* 7 (1), 35-59.

Browman, Catherine and Louis Goldstein. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18, 299-320.

Browman, Catherine and Louis Goldstein. 1992a. Articulatory Phonology: An overview. *Phonetica* 49, 155-180.

Browman, Catherine and Louis Goldstein. 1992b. "Targetless" schwa: an articulatory analysis. In G. Docherty and D. R. Ladd, Eds., *Papers in*

Laboratory Phonology II: Gesture, Segment, Prosody. Cambridge: Cambridge University Press.

Catford, J.C. 1988. *A Practical Introduction to Phonetics*. Oxford: Clarendon Press.

Davidson, Lisa. 2003. *The Atoms of Phonological Representation: Gestures, Coordination and Perceptual Features in Consonant Cluster Phonotactics*. Ph.D. dissertation, Johns Hopkins University.

Davidson, Lisa, Peter Jusczyk and Paul Smolensky. 2003. The initial and final states: Theoretical implications and experimental explorations of richness of the base. In R. Kager, W. Zonneveld and J. Pater, Eds., *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: CUP.

Elson, Benjamin. 1956. Sierra Popoluca syllable structure. *International Journal of American Linguistics* 13, 13-17.

Gafos, Adamantios. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20 (2), 269-337.

Gick, B. and I. Wilson. to appear. Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. In *Papers in Laboratory Phonology VIII*. Cambridge: Cambridge University Press.

Gick, Bryan. 2002. An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica* 59 (1), 38-48.

Hancin-Bhatt, Barbara and Rajesh Bhatt. 1998. Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19, 331-378.

Iskarous, Khalil. 1998. Vowel Dynamics and Vowel Phonology. In S. Kimary, S. Blake and E.-S. Kim, Eds., *The Proceedings of the Seventeenth West Coast Conference on Formal Linguistics*. Palo Alto: CSLI.

Li, Min, Chandra Kambhamettu and Maureen Stone. 2002. *Region based contour tracking for human tongue*. IEEE International Symposium on Biomedical Imaging: Macro to Nano (ISBI2002), Washington, DC, July 7-10, 2002.

Matteson, Esther and Kenneth Pike. 1958. Non-phonemic transition vocoids in Piro (Arawak). *Miscellanea Phonetica* 3, 22-30.

Price, P.J. 1980. Sonority and syllabicity: Acoustic correlates of perception. *Phonetica* 37, 327-343.

Smorodinsky, Iris. 2002. *Schwas with and without active control*. Ph.D. dissertation, Yale University.

Stone, Maureen. 1991. Imaging the tongue and vocal tract. *British Journal of Disorders of Communication* 26, 11-23.

Stone, Maureen. 1995. How the tongue takes advantage of the palate during speech. In F. Bell-Berti and L. Raphael, Eds., *Producing Speech: Contemporary Issues: A Festschrift for Katherine Safford Harris*. New York: American Institute of Physics, pp. 143-153.

Stone, Maureen and Edward P Davis. 1995. A head and transducer support system for making ultrasound iamges of tongue/jaw movement. *Journal of the Acoustical Society of America* 98 (6), 3107-3112.

Tarone, Elaine. 1987. Some influences on the syllable structure of interlanguage phonology. In G. Ioup and S. Weinberger, Eds., *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Cambridge: Newbury House Publishers.