

Addressing phonological questions with ultrasound

LISA DAVIDSON

Department of Linguistics, New York University, NY, USA

(Received 2 March 2004; accepted 28 September 2004)

Abstract

Ultrasound can be used to address unresolved questions in phonological theory. To date, some studies have shown that results from ultrasound imaging can shed light on how differences in phonological elements are implemented. Phenomena that have been investigated include transitional schwa, vowel coalescence, and transparent vowels. A study of consonant cluster phonotactics is presented as an example of how ultrasound methodology can be used to examine phonological issues. Five English speakers presented with phonotactically illegal non-words (e.g., /zgomu/) typically repaired these sequences with vowel insertion (e.g., [zəgomu]). Using ultrasound imaging, the production of these words is compared to legal sequences that are articulatorily similar, such as *succumb* and *scum* to assess the nature of the schwa found between /z/ and the following consonant. Results indicate that for some speakers, production of schwa in /zC/ sequences is not consistent with the phonological epenthesis of a schwa. Instead, speakers appear to be failing to sufficiently overlap the consonant gestures.

Keywords: *Ultrasound, phonology, phonotactics, tongue*

One goal of the study of phonology as a laboratory science is to examine the relationship between phonological representations and the details of phonetic implementation. By phonological form, researchers are usually referring to a composition of discrete sounds that is the output of the phonological grammar. Phonetic implementation is generally considered to be the gradient and variable speech stream, characterized by being continuous. It is often assumed that the phonological and phonetic components are modular, in that phonetic implementation takes the output of the phonological system as the “plan” on which the speech stream is based. However, this view has recently been challenged by studies which have shown that processes that have been treated as categorical in the phonological literature are much more complex, and often gradient in their application (Pierrehumbert, Beckman, & Ladd, forthcoming, and references therein). Consequently, it seems worthwhile to question whether the traditional phonological descriptions of various processes—like vowel transparency, schwa epenthesis, or palatalization, for example—really best explain the data in question.

Under the umbrella of laboratory phonology, a number of processes discussed in the phonological literature have been experimentally investigated using various methodologies (see Kingston & Beckman, 1990; and subsequent proceedings of the Laboratory Phonology conference). Many of these are articulatory studies that have investigated the hypothesis that the nature of phonological processes can be uncovered by examining articulatory coordination. For example, Zsiga (1995) uses electropalatography (EPG) to demonstrate that palatalization across word boundaries, as in “miss you”, applies optionally rather than across the board, unlike the word-internal process found in the related words “confess” and “confession”. In an examination of the effects of speech rate on the conditioning of phonological processes, Browman and Goldstein (1990) argue that alternations found in casual speech, including segmental insertions, deletions and assimilations can result from changes in magnitude and temporal duration in the articulatory gestures. On the basis of X-ray pellet data for American English, they demonstrate that the tongue tip raising of the apparently deleted [t] in the casual production of “perfect memory” as [prfɛkməmri] is still present, but it is completely overlapped by the labial closure gesture for the /m/, rendering it inaudible.

In addition to EPG, X-ray microbeam, and also electromagnetic midsagittal articulography (EMMA) (e.g., Perkell et al., 1992), ultrasound is a particularly useful technology for laboratory phonology. Ultrasound is a practical tool, since it is readily available (unlike EPG, which usually requires custom-made palates) and relatively inexpensive (about one-third the cost of EMMA). Since ultrasound does not allow for the tracking of particular flesh points on the tongue (as EMMA does), it is best suited to phonological questions which can be investigated by examining the shape of the whole tongue. In fact, this is one of the great benefits of ultrasound, since it allows researchers to visualize not just the vertical and horizontal movements of the points where EMMA pellets are placed, but it allows for the imaging of otherwise elusive elements such as the tongue root, or lateral movement when the tongue is imaged coronally (see Stone, this volume, for more on coronal transducer placement).

Researchers have already begun to use ultrasound to resolve questions that have not been satisfactorily answered by traditional phonological analysis. For example, Iskarous (1998) employed ultrasound to investigate how tongue trajectories in [i]–[a] and [a]–[i] sequences might correlate with the asymmetrical behavior of these sequences in vowel coalescence. In many languages that resolve hiatus (two sequential vowels in a word) through coalescence, [a] followed by [i] coalesces to [e], but [i] followed by [a] does not (Casali, 1998). Iskarous’s tongue shape data suggested that the trajectories of [a]–[i] and [e] have dynamic targets that are similar in direction, facilitating coalescence, whereas the contradictory trajectories of [i]–[a] and [e] discourage coalescence. Similarly, Gick and Wilson (forthcoming) used ultrasound to demonstrate that the percept of a schwa that occurs in English vowel+liquid sequences (e.g., heel [hi^əl]) arises as a solution to conflicting articulatory targets. Specifically, Gick and Wilson hypothesized that the acoustic schwa appears because the articulatory movements necessary to produce vowels and liquids may conflict: whereas the former have an anterior tongue root position, the latter are characterized by a retracted tongue root. Thus, the schwa is a by-product of the route taken to resolve this conflict.

Vowel transparency in languages such as Wolof and Hungarian is another phonological issue being investigated with ultrasound (Archangeli, Kennedy, Baker, & Racy, 2004; Benus, 2005). Hungarian has transparent vowels (TVs), which are usually high vowels that allow harmony between the root-initial and suffix vowels regardless of their own backness

values (e.g., Imi-hez, Imre(dim)-Allat. vs. Tomi-hoz, Tom(dim)-Allat.). Using ultrasound, Benus examines the effect of harmonic domain (front or back) on the articulation of the TVs. The ultrasound data show that the tongue body is significantly more retracted if the TV is in the back domain. Benus uses this evidence to argue that the phonological alternation between suffixes is dependent on the degree of articulatory backness in the transparent vowel preceding the suffix vowel.

One important issue that arises in the use of ultrasound is how to quantify the results. For example, Gick and Wilson (forthcoming) argue that a transitional schwa is perceived because the tongue passes through the position of canonical schwa when moving from a high vowel to a liquid. While this is an adequate descriptive account of the tongue trajectory, ultimately a statistical analysis is necessary to verify the results of a comparison of two conditions. This is particularly crucial for phonological analyses relying on experimental data, since the ability to claim that two productions are categorically different requires some kind of external verification.

The next section presents a preliminary study of the nature of transitional schwa during the production of non-native consonant clusters in English. In addition to discussing the theoretical motivation for the study, a method for measuring tongue shapes and a similarity metric and statistical procedure for comparing shapes are introduced. Acoustic information provides further information regarding the differences between transitional and lexical schwa.

The nature of the schwa in non-native consonant cluster production

Producing phonotactically illegal sequences

Studies of the acquisition of non-native phonotactic sequences by second language (L2) learners have often reported that speakers repair non-native consonant clusters by epenthesis of a vowel between the consonants (Tarone, 1987; Broselow & Finer, 1991). For example, Tarone (1987) reports that Korean speakers repair some illegal sequences by inserting [ə] (e.g., *class* → [kəlæs]). Similarly, English speakers producing foreign names often repair phonotactically impossible clusters by inserting [ə] (e.g., *Dvořak* → [dəvɔrʒək]).

However, results emerging from research within the Articulatory Phonology framework have raised the possibility that the vowel being produced between the consonants in the illegal sequence is not the insertion of a lexical vowel, but rather the failure to adequately overlap the consonant gestures (Browman & Goldstein, 1990; Gafos, 2002). In a language like English, consonants in a cluster are produced with a close transition, which means that the release of the first consonant in a C1C2 sequence is substantially overlapped by the target, or the constriction portion, of the second consonant (Catford, 1977). This is shown schematically in Figure 1.

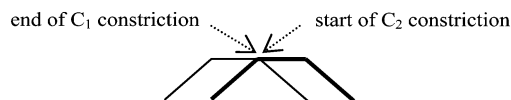


Figure 1. The lifespan of a consonant or vowel gesture is represented by the rising portion, or the onset of the gesture, the plateau, or constriction portion, and the falling portion, or the offset of the gesture. This schematic also demonstrates the overlapping coordination of two adjacent consonants.

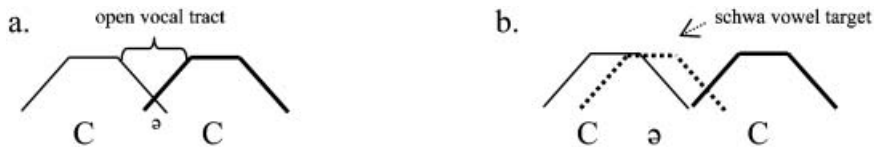


Figure 2. Gestural mistiming (a) versus schwa epenthesis (b).

In other languages, such as Moroccan Arabic (MA), consonant clusters contain open transitions, which are usually realized on the surface as a transitional vowel (e.g. /kteb/ → [kat^əb] “write (active participle)”) (Gafos, 2002). In MA, the presence of the transitional vocoid arises because the release of C1 is not overlapped by constriction portion of C2. Under these conditions, there is a period of open vocal tract where voicing can occur for a speaker in speech-ready state. This is shown schematically in Figure 2a. The case of the transitional vowel can be contrasted with a schwa that has its own gestural target, as in 2b.

The existence of different possibilities for consonant coordination suggest that L2 learners may have available to them multiple strategies for producing phonotactically impossible consonant clusters. When English speakers fail to produce sequences like word-initial /zg/ accurately, one option is to epenthesize a phonological vowel (as in 2b), or it might be that they are imposing a non-overlapping coordination on the consonants (as in 2a). Because speakers do not have experience with such clusters, it is plausible that they are unable to apply canonical English initial cluster coordination when attempting to produce them. The repair that gives rise to an transitional vowel (2a) can be called “mistiming”, while the insertion of a lexical schwa (2b) is referred to as epenthesis.

Because schwas from both mistiming and epenthesis are acoustically similar, an articulatory method such as ultrasound can help determine whether speakers repair non-native consonant clusters with the insertion of a schwa vowel, or by mistiming. To distinguish between epenthesis and 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) corresponding consonant sequences divided by a schwa. 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 then 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 transitional vowel, the tongue shape changes in [zəC] should be more similar to those produced for [sC]. These changes can be examined visually and statistically with ultrasound imaging.

Participants

The participants were five 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.

Experimental materials

The target stimuli were three triads of /sC-/ , /səC-/ , and /zC-/ initial words. The /zC/-initial words were possible but non-words in Polish, and all stimuli were recorded by a bilingual

English-Polish speaker using the Kay Elemetrics CSL at a 44.1-kHz sampling rate. Because it has been shown that voicing does not primarily affect the supraglottal constriction (e.g., House & Fairbanks, 1953; Perkell, 1969), the voiced and voiceless cognates of these clusters can be compared. An effort was also made to match all members of the triad on the vowel 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 I.¹ In addition to the target words, 24 more legal words and eight more non-words were recorded to present to the participants as fillers, for a total of 44 words.

Design and data collection

An ultrasound machine (Acoustic Imaging, Inc., Phoenix, AZ, USA, Model AI5200S) was used to collect midsagittal images of the tongue from the five 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 10 cm. The speaker's head was stabilized by a specially designed head and transducer support (HATS) system (for details, see Stone & Davis, 1995; Stone, 2005). The transducer was placed midsagittally under the speaker's chin so that the image of the tongue appears between the shadows of the jaw and the hyoid bone (for more on midsagittal transducer placement see Stone, 2005). The HATS system ensures that both the speaker's head and the transducer do not move during data collection, so that the images for individual stimuli can be directly compared.

The stimuli were presented to the speaker using PsyScope 1.2.6 on a Macintosh G3 laptop. The words appeared in random order 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.

Visualization

The acoustic record was examined to determine the time and duration of each /səC/, /sC/, and /zC/ sequence produced by the speakers. Since the audio and video streams are synchronized by the Canopus ADVC-1394 video acquisition hardware, the relevant ultrasound frames were chosen by calculating the frames corresponding to the acoustic start of the fricative /s/ or /z/ up to the burst of the following stop. Each sequence had a duration of approximately five frames (for the few repetitions that had a duration of six frames, the first five were used in the statistical analysis). The middle five of the seven repetitions of each sequence by the speaker were measured.

Table I. Target stimuli used in experiment.

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]

Tongue shapes are measured using EdgeTrak, an automatic system for the extraction and tracking of tongue contours from the ultrasound image (Li, Kambhamettu, & Stone, 2003; see also Li, Kambhamettu, & Stone, 2005). Once the tongue contours are tracked, they can be displayed as a series of x, y, t surfaces using the program SURFACES (Parthasarathy, Stone, & Prince, 2003, for more information on the advantages of SURFACES, see also Parthasarathy, Stone, & Prince, 2005).

Statistical measures

The measurement issues that arise in this study are different from those of previous ultrasound research because it involves determining whether two items are similar, not whether or not they are the same. For example, for the Hungarian transparent vowels discussed earlier, Benus (2005) isolated the ultrasound frames corresponding to [i] and chose the frame with the most extreme high front articulation for his comparisons. In this study, the objective is not to determine whether the production of [zəC] is the *same* as the production of [səC] or [sC], since these sequences do not contain exactly the same phonemes. Instead of matching up the frames corresponding to individual sounds, it was decided that the nature of schwa in a [zəC] sequence should be determined by examining the trajectory of the whole sequence over time, without emphasizing the exact phoneme that each frame represents. In other words, we can only be confident that the production of [zC] is more similar to either [səC] or [sC] if it can be shown that the trajectory over a period of 5 frames is consistently more similar to one or the other.

Another reason for examining the trajectory over several frames is that ultrasound has a relatively low sampling rate of 30 Hz. Because the duration of schwa hovers around 30 ms, it is possible that on any given repetition, the ultrasound machine did not record a frame during the production of the schwa. This makes it risky to assign a specific frame to schwa, but because it is hypothesized that a lexical schwa will coarticulate with surrounding gestures and affect the entire trajectory of the sequence,² an examination of the change over time gives a more global picture that allows us to assess whether [zəC] can be considered more similar to [sC] or [səC]. Furthermore, by collecting multiple repetitions, we can implement a statistical procedure that is more robust than if we attempted to compare the curvature of single productions.

There are three possible outcomes for the statistical procedure described below. First, for any given speaker and triad, the tongue shapes in the five frames corresponding to [zəC] could be more similar to those for [sC] significantly more often than those for [səC]. If this is true for several frames being compared, it would be consistent with the hypothesis that speakers are transitioning from the fricative to the following consonant without epenthesizing a lexical schwa between the consonants. Alternatively, if the frames for [zəC] are more similar to [səC] significantly more often, it would suggest that schwa insertion does make the articulatory trajectory of [zəC] comparable to [səC]. The third possibility is that the frames for [zəC] are not significantly more similar to either [səC] or [sC], or we find that a few frames are more similar to [sC], while others are more similar to [səC]. This third possibility is essentially a null result, since such patterns could arise either if the trajectory for [zəC] is equally different from [səC] or [sC], or if there are technical problems resulting from mismatched frames or a low sampling rate. In fact, given all of the potential sources of noise in this ultrasound study, null results for different speakers and triads would not be surprising, so a significant result is indicative of a strong similarity effect. Nevertheless, all of the potential problems with the ultrasound methodology will be kept in mind when interpreting the results.

The differences between the tongue shapes for the native and non-native sequences are quantified for each frame using an absolute value mean distance measure. This mean distance metric can be used to measure the differences between the curves for two frames based on the distance between the matrix representing one tongue shape and the matrix representing the other. Furthermore, this metric is sensitive not just to the shape of the curvature, but also to the location of the curves relative to one another. That is, the mean distance measure takes into account both the height and backness of the curves. The equation for the absolute value mean distance is shown in equation 1.³ To find the distances between the two curves, the curves are resampled at 100 points, and the distance is then computed between corresponding points on the curve. The distance equation is shown in equation 2. In equation 2, *a* and *b* refer to the tongue contours being compared. In both equations, *i* refers to an (x,y)-coordinate (1-100) on the tongue contours.

$$\hat{\mu}_{d, abs} = \frac{1}{N} \sum_{i=1}^N |d_i| \quad (1)$$

$$d_i = \sqrt{(a_{x,i} - b_{x,i})^2 + (a_{y,i} - b_{y,i})^2} \quad (2)$$

Once the mean distances between curves are established, the sign test is used to determine whether speakers' productions of [zəC] are statistically more similar to [səC] or [sC]. To create the input for the sign test, for each frame, the mean distance is calculated between every repetition of [səC] and every repetition of [zəC]. The same is done for every possible combination of repetitions for the [sC]-[zəC] comparison. This generates 25 mean distance values for both the [səC]-[zəC] comparison and the [sC]-[zəC] comparison for each frame, which is the input to a matched pairs sign test. If speakers' tongue shapes for [zəC] are reliably more like [sC], then the mean distance values for comparisons of the individual repetitions should be smaller for this pair significantly more often, regardless of which repetitions are compared.

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

Results: Articulatory data

To exemplify, spatio-temporal images produced with SURFACES are shown for speaker JED's velar triad (*succumb*, *scum*, *zgomu*) in Figure 3. Note that these figures do not represent one tongue surface, but rather are six frames that are connected to show how the tongue shape changes from frame to frame.

Impressionistically, a number of differences can be seen among the images in Figure 3. First, the initial two frames of [sək] in 3a, which roughly correspond to the [s], are 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 grey which corresponds to a lower position in [sək] and a darker grey which indicates a higher position for [sk] and [zəg] (3b and c). These plots suggest that when an [s] is followed by a [ə], it has a starting position lower in the oral tract because it is coarticulating with the [ə], which has a neutral (lower) tongue body

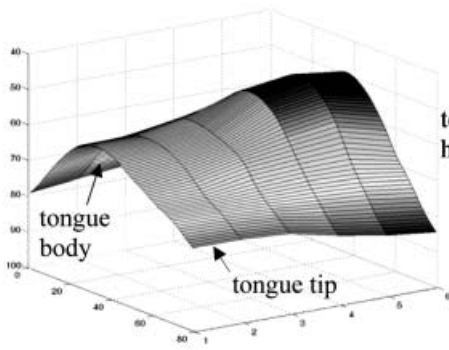
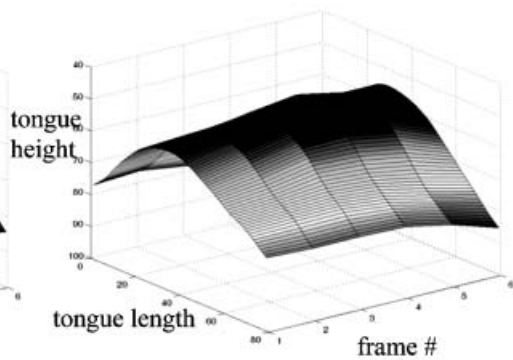
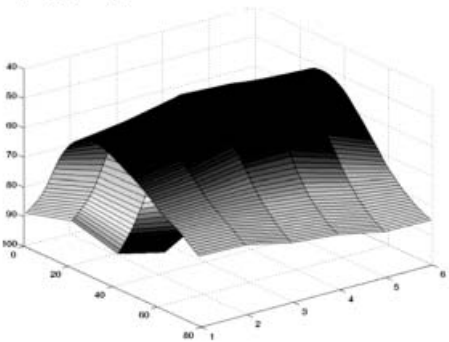
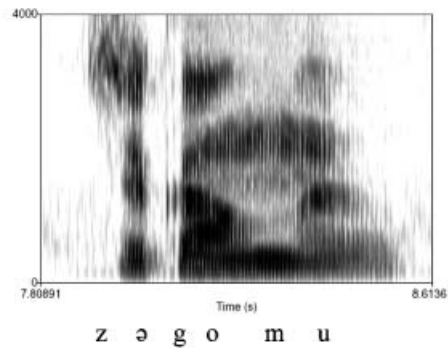
a. [sək] of *succumb*b. [sk] of *scum*c. [zəg] of *zgomu*d. spectrogram of JED's *zgomu*

Figure 3. The units of the x- and y-axes are millimetres; the origin is the top left of the ultrasound image. The shading reflects the height of the tongue curve: dark grey is a high position, and lighter grey is a lower position. In (d), the spectrogram for JED's utterance of [zəgomu] is shown.

position. When [s] is followed by [k], on the other hand, it is coarticulating with a gesture that has a high tongue body position, so the [s] itself is produced in a higher position. The [z] in [zəg] appears more like [sk], in that it also starts at a higher position.

Differences in the production of the non-native and native forms can also be visualized by superimposing each of the repetitions of successive frames for all of the words in a triad, as in Figure 4. This is shown for frames 2–4 of subject ELR's and HJC's labial triad (superfluous ~ spurt ~ zbura (ELR)/zbertu (HJC)). Acoustically, the /zb/ target sounds like [zəb] for both of these speakers. However, these figures show that for ELR, the dotted lines representing [sp] and the dashed lines representing [zəb] are more similar than the plain lines plotting [səp]. This pattern holds for all of the frames shown. For HJC, on the other hand, the opposite pattern holds, and [zəb] is more similar to [səp]. The difference in these patterns is supported by the quantitative data detailed below.

Each of the five speakers produced the coronal, labial, and velar triads. The final data set including all participants' utterances contained a total of five 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. Each speaker's production of stimuli with a /zC/ initial cluster is summarized in Table II.⁴

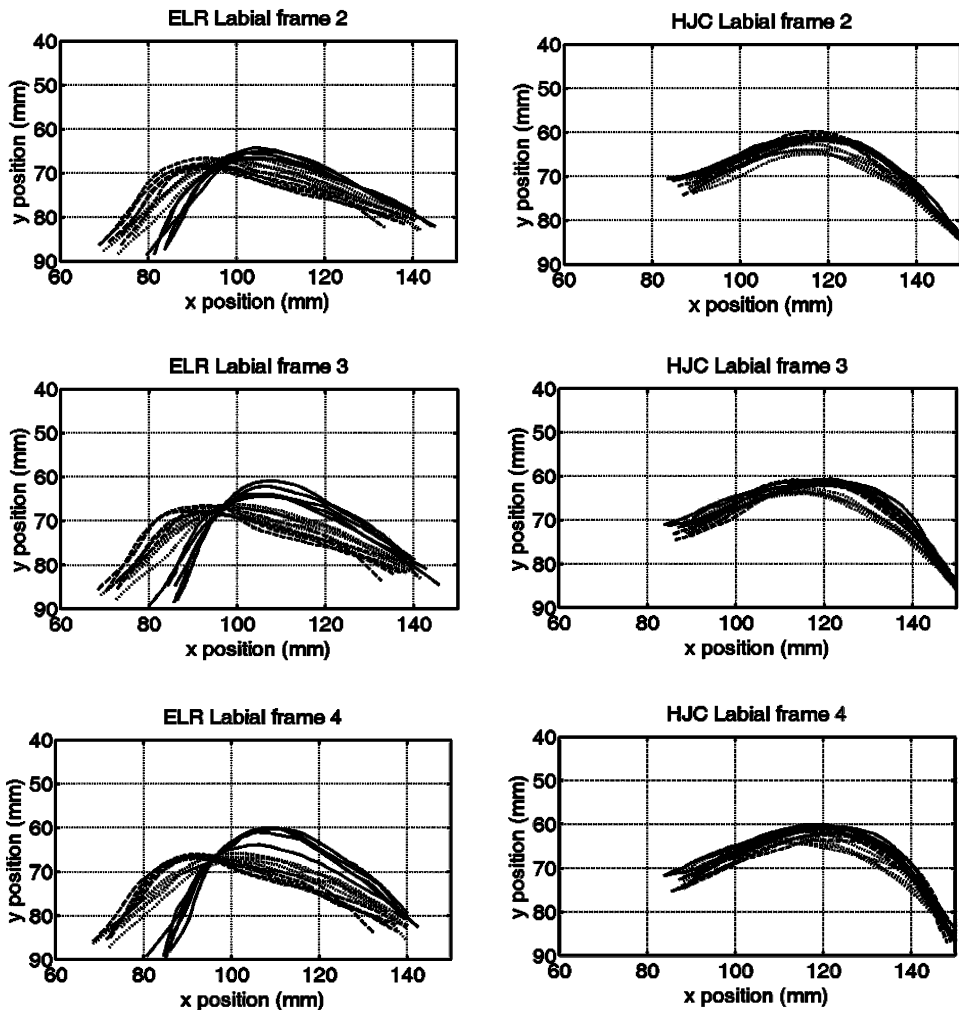


Figure 4. Comparison of tongue shapes for 3 example frames for speaker ELR's and HJC's production of [səp] (solid line), [sp] (light dotted line), and [zəb] (dark dashed line). All five repetitions of each word are shown on the plots. The x position corresponds to tongue length, and the y-position to tongue height, both in millimeters.

Mean distance results are reported for the 11 tokens in which a [ə] was produced between the consonants in the /zC/ cluster. For illustrative purposes, Table III shows the average of the 25 mean distance values that were used in the sign test. For each of the pairs of mean distance values, the smaller number represents a smaller average difference between the curves for that comparison and that frame.

These results for the cases when a schwa is present acoustically indicate two patterns. First, the [zəC] productions for ELR, JED, and PDD are more similar to [sC] than [səC]. This is supported by the fact that the majority of the differences between frames for these speakers are significantly smaller for the [zəC]–[sC] comparison. Second, while HJC produced a schwa in all contexts, her tongue shapes were more similar to [səC] for the labial case and the coronal case to some extent, but did not show statistically greater similarity to either pattern for the velar data. Likewise, the data for KAH is variable, and her productions do not seem to have a consistent pattern. As noted earlier, this data is

Table II. Speakers' productions of /zC/ initial clusters.

JED	HJC	PDD	KAH	ELR
[zderu]	[zədiri]	[zderu]	[zədiri]	[steru]
[zəbura]	[zəbertu]	[zəbura]	[zəbertu]	[zəbura]
[zəgomu]	[zəgomu]	[zəgomu]	[zəgomu]	[skama]

prone to noise from a number of different sources (i.e., mismatched frames, low sampling rate), so it is possible that more robust findings would result for KAH and HJC if a greater number of repetitions had been collected.

Results: Acoustic data

In addition to the articulatory data, acoustic data was also examined for both inserted and lexical schwa, including duration, F1 midpoint value, and F2 midpoint value. All measurements were taken using Praat (Boersma & Weenink, 2004). The duration of the vowel was measured from the onset of voicing or F2 (in voiced clusters) to the cessation of F1 and F2 at the obstruent closure. The midpoint F1 and F2 values were obtained using linear interpolation Burg LPC. Because there are only five tokens per speaker for each word, no statistics have been performed. Instead, the average schwa duration, F1 midpoint, and F2 midpoint values are reported for each speaker.

If the speakers are mistiming rather than epenthesizing, we expect to see at least two differences on the acoustic record. First, it is predicted that transitional schwa will generally

Table III. Mean distance results for [zəC] comparisons to [səC] (odd columns) and [sC] (even columns) for each of the 5 measured frames in the sequence. The mean distance values for significantly smaller differences are in bold (i.e., the sign test value is ≤ 7 when the [səC]–[zəC] comparison is significantly smaller, and ≥ 18 when the [sC]–[zəC] comparison is significantly smaller). italics indicate marginal significance (i.e., the sign test value is either 8 or 17).

Frames	KAH		HJC		PDD		JED		ELR		
	/səp/-/zb/	/sp/-/zb/	/səp/-/zb/	/sp/-/zb/	/səp/-/zb/	/sp/-/zb/	/səp/-/zb/	/sp/-/zb/	/səp/-/zb/	/sp/-/zb/	
LAB	1	1.58	1.78	1.29	1.70	1.67	1.45	1.71	1.51	4.06	1.85
	2	1.66	1.64	1.06	1.60	3.28	2.67	1.87	1.86	4.98	2.22
	3	1.67	1.54	1.22	1.61	4.24	2.78	2.66	2.29	6.38	2.35
	4	1.84	1.38	1.53	1.86	3.78	2.41	3.81	3.41	7.46	2.59
	5	1.80	1.35	1.96	2.07	2.50	2.21	4.80	4.27	7.35	2.96
		/sək/-/zg/		/sk/-/zg/		/sək/-/zg/		/sk/-/zg/		/sək/-/zg/	
VEL	1	1.71	2.09	1.84	1.98	5.10	4.71	2.71	1.86		
	2	2.13	2.23	1.98	1.81	6.15	4.97	4.21	2.98		
	3	2.70	2.58	2.33	2.38	6.99	5.69	4.87	3.32		
	4	3.19	2.86	2.70	2.68	7.55	6.08	2.79	<i>2.54</i>		
	5	2.53	2.45	2.27	1.98	7.72	6.07	2.98	3.81		
		/sət/-/zd/		/st/-/zd/		/sət/-/zd/		/st/-/zd/			
COR	1	1.54	1.77	1.45	1.52						
	2	1.57	1.97	1.25	1.52						
	3	1.54	2.12	1.33	1.55						
	4	<i>1.51</i>	2.16	1.55	1.57						
	5	<i>1.64</i>	1.99	2.05	1.98						

be shorter than lexical schwa, since the release of C1 and onset of C2 may still be partially overlapped even when an transitional vowel is present (as demonstrated in Figure 2a). Second, it is expected that F1 will be lower for a transitional schwa. Since a transitional schwa is a brief period of open vocal tract between two constrictions (e.g., from fricatives to stops), the mouth is likely to be more closed than it is for a lexical schwa, leading to lower F1 (Kondo, 1994; Flemming, 2004). The behavior of F2 is less clear, since it is well known that F2 of schwa varies greatly depending on consonantal context and the following vowel. Because this study could not strictly control for the vowel following the cluster both within and across the triads, no predictions regarding F2 are investigated here, though the data is still reported Table IV. Mean schwa duration and F1 and F2 midpoint values for each speaker and each triad are shown in (see Davidson, forthcoming, for an extended acoustic study of transitional schwa containing discussion of the role of F2).

The acoustic data indicate that seven of the 11 average schwa durations are shorter for [zəC] than [səC], whereas the remaining durations are approximately the same (PDD velar, KAH velar and coronal, HJC coronal). These findings generally match up to the articulatory results that suggest that ELR, JED, and PDD are more likely to be mistiming than KAH and HJC.

The F1 midpoint data indicates that for all of the velar and coronal tokens, F1 for [zəC] is considerably lower than [səC]. For the labial tokens, two of the five [zəb] tokens had a lower F1, whereas the other three were either the same or slightly higher. One possible reason for this discrepancy is that the vocal tract resonances both for the lexical schwa in *superfluous* and the transitional schwa in [zəb] might be affected by the following labial stop and [ɣ] for some speakers. Both of these phonemes cause considerable lip rounding, which may affect the shape of the vocal tract from the beginning of the utterance and which could potentially neutralize any F1 distinctions that might arise between the lexical and transitional schwas. Specifically, the vowel with the expected higher F1—the lexical schwa—is more likely to be slightly lowered as a result of lip rounding, whereas sounds with an already lower F1 will be less affected by the rounding (Stevens, 1998).

Table IV. Mean durations for schwa (in milliseconds) and mean F1 and F2 midpoint values (in Hz) for [zəC] and [səC].

Measure	KAH		HJC		PDD		JED		ELR	
	/səp/	/zəb/	/səp/	/zəb/	/səp/	/zəb/	/səp/	/zəb/	/səp/	/zəb/
LAB schwa dur.	46	34	56	34	36	30	40	25	61	56
F1 midpoint	541	443	378	371	295	332	361	283	402	567
F2 midpoint	1541	1537	1729	1595	1640	1548	1431	1437	1642	1460
	/sək/	/zəg/	/sək/	/zəg/	/sək/	/zəg/	/sək/	/zəg/		
VEL schwa	40	43	61	53	36	36	43	35		
F1	575	524	398	337	387	304	370	310		
F2	1551	1307	1563	1347	1541	1743	1644	1538		
	/sət/	/zəd/	/sət/	/zəd/						
COR schwa	54	68	64	63						
F1	461	412	401	347						
F2	1659	1727	1561	1732						

General discussion

It was hypothesized that greater similarity of [zəC] to [sC] tongue shapes would arise if the phonological output for /zC/ word-initial clusters does not include a schwa gesture with its own target. For at least three speakers (ELR, PDD, JED), the trajectory of the tongue over time during the production of [zəC] is more similar to the trajectory for [sC] for a majority of the frames, which is consistent with the hypothesis that speakers are failing to produce the [z] and subsequent consonant with sufficient overlap. If the constriction of the vocal tract is even slightly released between the two consonants, a vowel will be perceived (Gafos, 2002). If a phonological 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.

For these three speakers, the acoustic results are consistent with what is expected for a transitional schwa: shorter durations since C1 and C2 should be somewhat overlapped, and lower F1 since the speakers are moving from one constriction to another, which results in a more closed oral tract. If PDD is indeed mistiming the gestures, then his velar triad is an interesting case showing that there is variation in the amount of overlap that the speakers impose on the consonants. That is, the schwa in his velar [zəg] is the same duration as [sək], but [zəb] is slightly shorter than [səp]. Since English speakers have no experience with coordinating these consonants into a cluster, they are not expected to coordinate them the same way every time. In general, though, the transitional schwas tend to be shorter, suggesting that the period of open vocal tract for transitional schwa has a shorter duration than lexical schwa.

The patterns observed for these three speakers are also similar to those found in a large-scale acoustic study of lexical versus transitional schwa (Davidson, forthcoming). In that study, speakers produced both CC clusters and matching CəC sequences in 18 different consonantal contexts, such as *zgadi/zəgadi* and *ftalu/fətalu*. When speakers repaired a cluster with a transitional schwa, it was compared to the lexical schwa in the same context. Results showed that the transitional schwa was significantly shorter and its F1 was significantly lower, regardless of consonantal context. Whereas it might be expected that F1 for some kind of inserted schwa will vary in different environments, it would not be the case that F1 would be lower in every environment unless it reflects a more closed mouth for transitional schwa than for lexical schwa.

The remaining two speakers do not appear to be using a uniform approach to produce the /zC/ targets. The articulatory data shows that [zəC] is neither consistently more similar to [səC] or [sC], and the acoustic data is similarly difficult to interpret. In some cases, the [zəC] schwas are equal in length or even longer than those in [səC], though those same schwas have lower F1 values than the lexical schwa, which is more likely a property of transitional schwa. 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 perhaps an entirely different strategy that has not yet been considered.⁵ Alternatively, it is possible that there is not enough data for these speakers. Note that apparent differences in repair strategies across speakers is not problematic, since there is no expectation that all speakers must exhibit the same strategy for producing non-native consonant clusters. The phonology gives the option of many different kinds of repairs and it is expected that any or all of these options could be observed.

Conclusions

This paper has detailed some studies that have productively used ultrasound to examine how articulatory information can inform unresolved questions in phonological theory. This line of work has followed from the notion prevalent in the laboratory phonology literature that not all phenomena previously described by phonological rules are necessarily abstract, categorical processes that apply across the board. In some cases, such information has suggested that mental-level phonological explanations are not necessary to account for the production facts (transitional schwa preceding English liquids, Gick & Wilson, forthcoming). In other cases, however, articulatory evidence has provided an alternative explanation for a puzzling phonological phenomenon (transparent vowels in affix selection in Hungarian, Benus, 2005).

The preliminary study on the nature of schwa has provided tentative evidence that the presence of vowel-like material between the two consonants of a non-native sequence does not necessarily represent phonological epenthesis, but may instead reflect speakers' failure to adequately overlap the gestures in that sequence. The evidence provided from ultrasound imaging raises the possibility that not only are traditional repairs like epenthesis available to speakers, but that temporal factors may also affect production of non-native sequences. The possibility of such a repair is interesting because it has implications for the developmental process of L2 acquisition, especially regarding the types of knowledge that must be acquired by L2 learners.

In addition to the theoretical findings of this study, it was also argued that research using ultrasound requires adequate measurement and statistical techniques in order to interpret the data. The measurement applied to assess the similarity of a target utterance (here, /zC/ clusters) to two other patterns (/sC/ or /səC/) was the absolute value of the mean distance, which provides a value for the difference between two tongue curves. These values were then submitted to a sign test, a conservative statistical test. The development of these and other metrics designed to interpret ultrasound images will make ultrasound an even more promising tool for both resolving old questions in phonology and asking new ones.

Acknowledgements

Thanks to Maureen Stone, Stefan Benus, David Goldberg, Christine Mooshammer and Amanda Miller-Ockhuizen for their insightful comments and/or help with the data. Parts of this work were presented at WCCFL 22 and submitted as part of the author's doctoral dissertation (Johns Hopkins University, 2003). This research was supported by the IGERT program in the Cognitive Science of Language at Johns Hopkins University, National Science Foundation Grant 997280 and by the National Institutes of Health Grant DC01758 to Dr. Maureen Stone.

Notes

1. Because *superfluous* has alternate pronunciations of either [sup]erfluous or [səp]erfluous, only speakers who clearly produced the schwa variant were recorded.
2. Attempts to demonstrate that schwa is not a segment but an interpolation between the surrounding segments have been forced to conclude that schwa does have a gestural target (Browman & Goldstein, 1992; Kondo, 1994). For example, Kondo's acoustic data indicate that while schwa might not be specified for backness (i.e., F2 is affected by the surrounding environment), it does have a consistent F1 value indicating that schwa does have a target height.

3. Another possible measurement is the root mean square, which is very similar to the mean distance measure in equation 1, except that it uses squared values. This amplifies the larger distances and reduces the influence of the smaller ones. In general, the two measures return very similar results, but the absolute value of the distance is chosen as being a more direct measure.
4. Although each speaker produced two tokens for each of the /zC/ targets, only the stimulus with the best image for each speaker was measured.
5. One reviewer suggests that an alternative explanation for all of the speakers' behavior is that they are producing lexical schwas, but that they are undershooting the targets. The theory of undershoot was proposed by Lindblom (1963), who demonstrated that as Swedish vowels decrease in duration, they also tended to fail to reach their target formant frequencies. However, subsequent studies failed to show similar effects in other languages, including American English (Gay, 1978; Fourakis, 1991) and Dutch (van Son and Pols, 1992). Moon and Lindblom (1994) addressed these critiques by demonstrating that undershoot occurs when there is a large difference between the tongue position for C1 and the tongue position for the following vowel (as in the [wi] of *will* and *willing*). In this study, all C1s are coronal fricatives, and all following vowels are schwas. Even if some speakers are inserting schwa, this combination of sounds, in addition to the short duration of schwa, is very unlikely to lead to undershoot.

References

- Archangeli, D., Kennedy, R., Baker, A., & Racy, S. (2004). *Ultrasound techniques for phonological research*. Department of Linguistics Colloquium: University of Arizona.
- Benus, S. (2005). Dynamics and Transparency in Vowel Harmony. PhD dissertation. USA: New York University.
- Boersma, P., & Weenink, D. (2004). *Praat: Doing phonetics by computer (v. 4.2)*. Computer program, available at: <http://www.praat.org>.
- Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research*, 7, 35–59.
- 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. (1992). "Targetless" schwa: an articulatory analysis. In G. Docherty, & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press.
- Catford, J. C. (1977). *Fundamental Problems in Phonetics*. Bloomington, IN: Indiana University Press.
- Casali, R. (1998). Resolving Hiatus. New York and London: Garland Publishing.
- Davidson, L. (forthcoming) Phonology, phonetics, or frequency influences on the production of non-native sequences. *Journal of Phonetics*.
- Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *The Phonetic Bases of Markedness*. Cambridge: Cambridge University Press.
- Fourakis, M. (1991). Tempo, stress, and vowel reduction on tongue position in American English. *Journal of the Acoustical Society of America*, 90, 1816–1827.
- Gafos, A. (2002). A grammar of gestural coordination. *Natural Language and Linguistic Theory*, 20, 269–337.
- Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America*, 63, 223–230.
- Gick, B., & Wilson, I. (forthcoming). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. In, *Papers in Laboratory Phonology VIII*. Cambridge: Cambridge University Press.
- House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25, 105–113.
- Iskarous, K. (1998). Vowel Dynamics and Vowel Phonology. In S. Kimary, S. Blake, & E.-S. Kim (Eds.), *The Proceedings of the Seventeenth West Coast Conference on Formal Linguistics*. Palo Alto, CA: CSLI.
- Kingston, J., & Beckman, M. (Eds.) (1990). *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press.
- 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). Edinburgh: Theoretical and Applied Linguistics Department, University of Edinburgh.
- Li, M., Kambhamettu, C., & Stone, M. (2003). Snake for band edge extraction and Its applications. Paper presented at 6th IASTED International Conference on Computers, Graphics, and Imaging, August, Honolulu, HI, USA.

- Li, M., Kambhmettu, C., & Stone, M. (2005). Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics*, 19, 545–554.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773–1781.
- Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, 96, 40–55.
- Parthasarathy, V., Stone, M., & Prince, J. (2003). Spatiotemporal visualization of the tongue surface using ultrasound and kriging (surfaces). In R. Galloway, Jr. (Ed.), *Proceedings of SPIE-Medical Imaging*, vol 5029. Bellingham, WA: International Society for Optical Engineering.
- Parthasarathy, V., Stone, M., & Prince, J. (2005). Spatiotemporal visualization of the tongue surface using ultrasound and kriging (surfaces). *Clinical Linguistics and Phonetics*, 19, 529–544.
- Perkell, J. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, MA: MIT Press.
- Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., & Jackson, M. (1992). Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America*, 92, 3078–3096.
- Pierrehumbert, J., Beckman, M., & Ladd, D. R. (forthcoming). Conceptual foundations of phonology as a laboratory science. In N. Burton-Robert, P. Carr, & G. Docherty (Eds.), *Phonological Knowledge: Its Nature and Status*. Cambridge: Cambridge University Press.
- Stevens, K. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19, 455–501.
- Stone, M., & Davis, E. P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of the Acoustical Society of America*, 98, 3107–3112.
- 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.
- 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.
- Zsiga, E. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In B. Connell, & A. Arvaniti (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 282–302). Cambridge: Cambridge University Press.