Original Paper

Phonetica 2004;61:220–233
DOI: 10.1159/000084159

Received: February 20, 2004
Accepted: November 10, 2004

Language-Specific Articulatory Settings: Evidence from Inter-Utterance Rest Position

Bryan Gick[a, b] Ian Wilson[a] Karsten Koch[a] Clare Cook[a]

[a]Department of Linguistics, University of British Columbia, Vancouver, B.C., Canada;
[b]Haskins Laboratories, New Haven, Conn., USA

Abstract

The possible existence of language-specific articulatory settings (underlying or default articulator positions) has long been discussed, but these have proven elusive to direct measurement. This paper presents two experiments using X-ray data of 5 English and 5 French subjects linking articulatory setting to speech rest position, which is measurable without segmental interference. Results of the first experiment show that speech rest position is significantly different across languages at 5 measurement locations in the vocal tract, and is similar to previously described language-specific articulatory settings. The second experiment shows that the accuracy of achievement of speech rest position is similar to that of a specified vowel target (/i/). These results have implications for the phonetics and phonology of neutral vowels, segmental inventories, and L2 acquisition.

Introduction

That languages differ in their general pronunciation tendencies has been noted by scholars since at least the 7th century AD [Laver, 1978]. Sweet [1890, p. 69] called such a tendency the 'organic basis' of a language, and he stated: 'Every language has certain tendencies which control its organic movements and positions, constituting its organic basis or the basis of articulation. A knowledge of the organic basis is a great help in acquiring the pronunciation of a language.' Honikman [1964] revived the study of organic basis in the English literature, and she gave it a new name: 'articulatory setting' (AS). She defined AS as the 'gross oral posture and mechanics' required for the 'economic and fluent' production of the 'established pronunciation of a language' [Honikman, 1964, p. 73] (Laver [1978] and Jenner [2001] give excellent historical surveys of this concept). Indeed, if such a postural basis underlies every language, this must not only contribute in large part to the overall 'sound' of a language or dialect, but must also interact with its phonetic inventory and phonological patterns – both influencing and being influenced by them – in as yet unknown ways, and must greatly influence how this overall sound changes over time.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/04/0614–0220
$21.00/0
Accessible online at:
www.karger.com/journals/pho

Dr. Bryan Gick
University of British Columbia
Department of Linguistics
1866 Main Mall, Buchanan E270
Vancouver, BC V6T 1Z1 (Canada)
Tel. +1 604 822 4817, Fax +1 604 822 9687
E-Mail gick@interchange.ubc.ca

Despite its prominence in the literature, AS has proven very elusive to direct measurement. In the past, this was because necessary measurement techniques did not exist. Heffner [1950, p. 99] says: 'No method of measurement has been devised that would permit the mathematical description of a basis of articulation.' O'Connor [1973, p. 289] calls for future studies of 'bases of articulation' and says: 'We know a good deal more about the detailed articulatory movements in a language than we know about the general articulatory background on which they are superimposed.' More recently, Collins and Mees [1995, p. 422] point out that 'at the moment, much of the description of AS features – including our own – is largely impressionistic'. A confounding reason behind the lack of quantitative evidence for ASs is the fact that 'no articulatory setting normally applies to every single segment a speaker utters' [Laver, 1978, p. 11]. In other words, segmental context has an overriding effect on the position of the articulators at any given time, making it difficult, if not impossible, to analyze speech with the aim of ascertaining the underlying AS of a given language.

Previous attempts to measure AS directly have focused primarily on identifying general acoustic properties of one language versus another, using broad measures such as long-term average spectra (LTAS). Harmegnies and Landercy [1985], for example, find small differences in LTAS between Dutch and French, but conclude that these are most likely based on differences in the segmental inventory. Similarly, Byrne et al. [1994] and Parry et al. [2000] find significant differences in the spectral envelope of speech across languages. Approaching the question from a different angle, Disner [1980] and Bradlow [1995] show that the same vowel category may be realized in different locations in the vowel space in different languages. While the results of all of these studies are consistent with the notion of underlying language-specific ASs, as Laver [1978] points out, one can never be sure whether acoustic differences in the speech signal are the result of differences in AS or differences in segmental targets. Further, while there may indeed exist specific categories of vocalizations in a language that reflect the properties of language-specific ASs (such as 'neutral' vowels in schwa or hesitation pause utterances), it remains unclear whether these types of utterances have their own targets [Browman and Goldstein, 1992; Gick, 2002a; Clark and Fox Tree, 2002]. These cases are discussed further later in this paper. Thus, while it may be possible for either the language learner or the analyst to recover information related to AS from various aspects of the acoustic signal, we cannot be certain of this until we are able to filter out the effects of targeted sounds.

An alternative means of testing for the existence of AS-like default positions for speech may be through language-specific rest positions during pauses between speech utterances. Öhman [1967] gives the earliest quantitative evidence we can find for the existence of a speech rest position (as distinct from an absolute rest posture), at least at the muscular level, for some articulators. Öhman's [1967, p. 43] EMG data show that 'the subject prepares himself for speech by certain postural adjustments' and that 'the articulatory movements of speech are modulations superimposed upon this basic posture' [p. 34]. In a study based on nonsense-word X-ray speech data pronounced by speakers of American English, Perkell [1969] uses the term *pre-speech posture* for this position. He finds that the larynx, the velum, and the tongue each take up a consistent posture when a speaker is preparing to speak. More recently, Barry [1992] mentions the existence of a speech posture as distinct from absolute rest posture, and Gick [2002a] provides X-ray evidence supporting this view, also for American English. While these studies lend support to the existence of a consistent speech-specific rest position within a particular speaker, they do not show whether this speech rest position is shared across speakers of a given language, whether speech rest positions are the same or different across languages, nor whether they are active during, and exercise influence on, speech itself. It therefore remains unclear whether or not these previous results are relevant to the question of language-specific ASs.

Some models of speech production propose using one or multiple equilibrium (or 'rest') positions, somewhat analogous to ASs as described here, as control parameters. According to Kelso et al. [1985], the equilibrium position of a mass-spring system is defined as the position where the net force on the mass is zero. In this model, the speech rest position that articulators assume between utterances must also be such a specified equilibrium position. Because they observe an inverse relationship between the stiffness of a specific articulator and its mean displacement from rest, they infer an equilibrium position as a 'key control parameter' [Kelso et al., 1985; Vatikiotis-Bateson, 1988], suggesting that these positions are maintained throughout utterances, not only active between them. Vatikiotis-Bateson [1988] applies this model to the analysis of speech production in French, English and Japanese, allowing for a cross-linguistic comparison of languages with different speaking rates and temporal organization. He found that the model developed in Kelso et al. [1985] accounted for kinematics in each language, but that there were still cross-linguistic timing differences between languages, suggesting language-specific settings of the two parameters stiffness and equilibrium position. In the equilibrium point hypothesis of Perrier et al. [1996], the equilibrium point is changed by central neural commands that produce a new (threshold) target muscle length. Differences between the actual and the threshold muscle lengths can produce movement to a new equilibrium position. Whether or not they are reset in this way, if there exist language-specific threshold muscle lengths for speech rest positions, then these models predict that the speech rest position should function (and be represented) as any other articulatory target.

If language-specific ASs do indeed exist, it becomes relevant to ask whether they are specified parts of a language's inventory (and hence learned from other speakers) or functionally derived properties of speech motor production. An AS of the first type, learned as a specific target, might either be maintained throughout an utterance, and hence be available to learners continuously through the acoustic signal (in which case the AS may be detectable by LTAS-type acoustic methods) or it may only be used as a language-specific rest position between utterances, requiring learners to pick up on audible cues at the starting and finishing transitions of utterances. In either case, however, if AS is an independent target, it need not be closely related to other elements in the phonetic inventory. In the case of AS being functional, at least two possibilities exist. First, AS could naturally arise out of motor efficiency requirements, and thus be relatable to the token frequency of articulatory targets in the language. In this scenario, for example, a language having a high frequency of postvelar sounds should be expected to exhibit a more retracted AS, thus decreasing the average 'travel cost' [Rosenbaum et al., 1995, 2001] of the tongue. A second functional possibility is that AS could be derived from type frequency of articulatory targets in the language. This is similar to the token frequency scenario in that the need for motor efficiency is being accommodated, but the efficiency in this case would be determined according to an awareness of the language's phonetic inventory, and hence could not be considered as having a purely motor basis. While the goal of the present paper is limited to determining whether language-specific ASs exist, bearing in mind these questions of the nature and origin of AS will help both in interpreting the results and in formulating questions for further research.

In the present paper, experiment 1 tests the hypothesis that speech rest positions are language-dependent, and that they relate directly to ASs in English and French. Upper lip, lower lip, pharynx, velum, tongue and jaw positions are measured during inter-utterance speech rest positions (henceforth 'inter-speech posture', ISP). A significant difference in these positions between the speakers of the two languages would provide quantitative evidence that a different baseline articulatory posture exists for each language. This posture can then be compared with previous descriptions of ASs, neutral vowels, and the like.

Experiment 2 tests the hypothesis that rest positions are specified in a manner similar to actual speech targets, as predicted by Perrier et al. [1996] and others (see above), by comparing the accuracy (standard deviation) of the tongue's return to ISP with the same accuracy in reaching a specified vowel target, /i/.

Experiment 1

In this experiment, X-ray films of English and French speakers were examined during inter-utterance rest position, and measurements were taken at seven locations in the vocal tract, to test whether AS is language-specific.

Methods

The subjects, shown in table 1, were the 5 English speakers (3 female, 2 male) and 5 French speakers (3 female, 2 male) from Munhall et al.'s [1994] Queen's University/ATR X-ray database. All French speakers were from Quebec. All English speakers were from the three Canadian provinces of British Columbia, Ontario, and New Brunswick. Each speaker in the database read between 25 and 30 sentences at a normal speech rate. It should be noted that the recording of subject 1 was slightly different than those of the other 9 subjects; the subject was recorded saying predominantly words, rather than sentences. In all but one of the tokens, the previous utterance was a single word, not a sentence. In addition, for subject 1 the X-ray was taken from a slightly different angle.

For all subjects, target frames at the midpoint of ISPs were selected from the database in digital form using Adobe Premiere, and the frames were exported as 360 × 240 PICT files. Individual frames were extracted at the midpoint of inter-utterance pauses (brief pauses between otherwise continuous sequences of sentences or single-word utterances in the database). These pauses ranged in duration from a minimum of 3 frames to a maximum of 16 frames (i.e. approximately 100–533 ms) and they corresponded to the period when the articulators had finished moving after articulating the previous sentence/word but before the articulators had started moving to articulate the following sentence/word. Pauses during which the subject swallowed, licked his/her lips, or performed some other active non-speech act were eliminated, leaving a minimum of 15 tokens (out of 25–30) for all but 2 subjects. The chosen individual frames were then imported into NIH Image (<http://rsb.info.nih.gov/nih-image/download.html>) for analysis. All measurements were made using the linear measurement tool. As there was no external point of reference available in the X-ray images from which to determine a scale, all measurements are given in pixels. Other than for subject 1, a male English speaker who contributed 7 of the 68 English tokens used, all subjects were in the same position and distance away from the camera, and thus a pixel would be the same real size for all other speakers [Rochette, 1973, vol. 1, p. 30, vol. 2, pp. 11–12]. Measurement reliability was tested by having one of the experimenters redo half of the measurements months after the original measurements had been made. The two sets of measurements matched each other very closely, always within 2 pixels and usually within 1 pixel. The longitudinal positions of the measurement points were tightly constrained within each subject, remaining within a 5-pixel range. Between subjects, the longitudinal positions varied somewhat, but this is to be expected given the differences in vocal tract morphology.

Table 1. Subjects and number of tokens

Subject	Gender	Age	Language	Film(s)	Tokens used, n
1	male	38	English	MIT–KNS	7
2	male	26	English	77	15
3	female	20	English	73, 74	16
4	female	22	English	80, 81	15
5	female	25	English	78, 79	15
6	female	20	French	24, 25	20
7	female	22	French	30, 31	11
8	female	29	French	26, 27	15
9	male	23	French	48, 49	19
10	male	30	French	32, 33	15

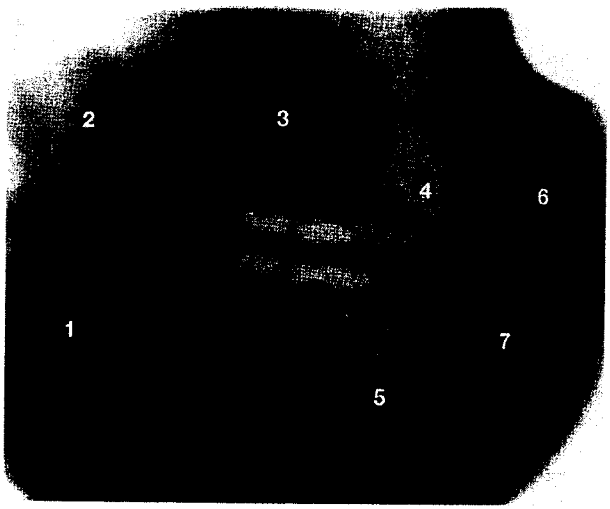

Fig. 1. The measurements taken were: 1 = pharynx width; 2 = velopharyngeal port width; 3 = tongue body distance from hard palate; 4 = tongue tip distance from alveolar ridge; 5 = lower-to-upper jaw distance; 6, 7 = upper and lower lip protrusion (distance from central incisors).

Measurements that were taken in experiment 1 included: (1) pharynx width: minimum distance from back of tongue to rear pharyngeal wall; (2) velopharyngeal port: minimum distance from velum to rear pharyngeal wall; (3) tongue body: minimum distance from tongue body to hard palate; (4) tongue tip: minimum distance from tongue tip to alveolar ridge; (5) jaw aperture: distance from lower jaw (chin) to upper jaw (upper central incisors); (6) u-lip protrusion: protrusion of upper lip from upper central incisors, and (7) l-lip protrusion: protrusion of lower lip from lower central incisors (fig. 1).

Statistical analysis was performed with ANOVAs, calculated in StatView 5.0.1. Raw measurements for each subject were normalized based on the subject's jaw size. As mandible length is correlated with sex [Rosas et al., 2002], this method can therefore help to minimize error across the small subject pool. Further, body size has also been correlated with other measures such as vocal tract length [Fitch and Giedd, 1999], and palate width and depth have also been used to predict race and sex significantly above chance [Burris and Harris, 1998]. Thus, as we have no data on subjects' body sizes, assuming a higher correlation between jaw size and the rest of the body than across subjects, normalizing to jaw size will be likely to decrease rather than increase across-subject variation. To calculate the normalization factor, subjects' jaw sizes were measured, and the average was given a value of 1.00. Jaw size for each subject was calculated by summing two distances: the distance from the superior

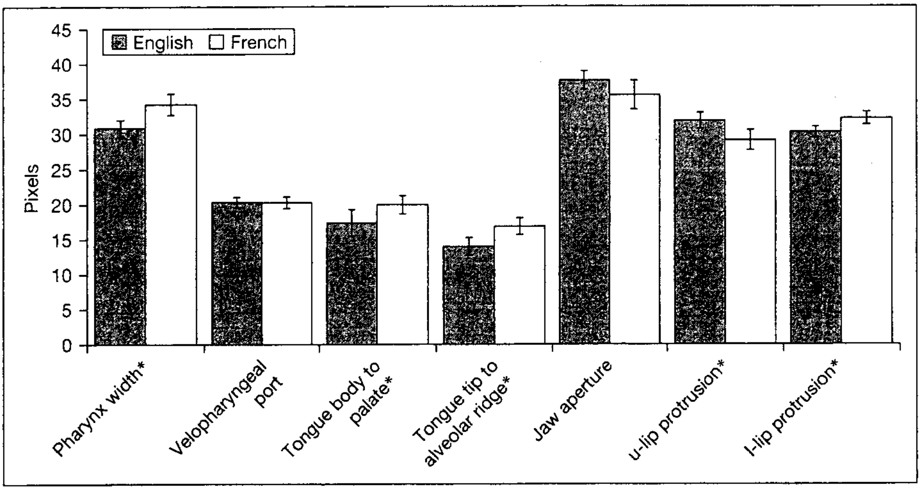

Fig. 2. Plot comparing English versus French mean measurements at 7 locations. Error bars show 95% confidence intervals. Significant differences between English and French were found at 5 of the locations, indicated by asterisks. For the purpose of clarity of display, jaw aperture values shown here are 60 pixels less than their actual values.

edge of the lower central incisors to the inferior edge of the mental protuberance plus the distance from the mental protuberance to the angle of the mandible. Each individual subject's jaw size was then normalized relative to the subject-wide average.

Phonetic environment was controlled for somewhat by eliminating all French and English tokens with a preceding rounded vowel, and also by maintaining a balance of tokens with and without a following rounded vowel or consonant. Because of limitations on the size of the database, it was not possible to balance the data for all factors in the experiment. Cases did exist in which the final sound preceding the inter-utterance pause was a sound in French that does not occur in English (e.g. nasalized vowels). In future studies where stimuli may be more carefully tailored, it will be desirable to control for environments that may affect the measurements taken in (1) to (5), in particular those that may affect tongue position in the palatal and pharyngeal regions.

Results: French versus English Rest Positions

ANOVA results revealed a significant difference between French and English speakers for five of the seven measurements. Results can be seen in figure 2. Compared with English, French had a significantly greater pharynx width [$F(1, 145) = 11.592$, $p = 0.0009$], a significantly lower tongue body [$F(1, 145) = 4.932, p = 0.0279$], a significantly lower tongue tip [$F(1, 130) = 10.489, p = 0.0015$], a significantly less protruded upper lip [$F(1, 108) = 8.525, p = 0.0043$], and a significantly more protruded lower lip [$F(1, 108) = 9.473, p = 0.0026$]. The two measurements that were not significant were velopharyngeal port width ($p = 0.9375$) and jaw aperture ($p = 0.0901$).

Because of the small number of subjects in the dataset, the jackknife procedure was applied to validate the above results. Means for all possible subsets of 4 speakers of each language were calculated. The resulting means (given in table 2) cluster together, and their distributions differ in the same way that the means calculated from all speakers' data do, suggesting that these results are likely to be representative of the larger (French and English) populations.

Table 2. Ranges of means from jackknife procedure for each vocal tract measurement (excluding those not found to be significantly different using ANOVA)

	English	French
Pharynx width	29.757–32.378	32.567–36.576
Tongue tip to alveolar ridge	12.726–15.447	15.245–17.984
Tongue body to palate	14.832–19.383	19.275–21.428
Upper lip protrusion	30.287–33.581	27.097–31.308
Lower lip protrusion	29.656–31.344	30.511–33.509

Discussion of Experiment 1

As predicted, the results of experiment 1 show pervasive differences in inter-utterance rest position between French and English, with French speakers exhibiting significantly greater lower lip protrusion, pharynx width and tongue tip to alveolar ridge distances ($p = 0.0015$), while English speakers had significantly greater upper lip protrusion and tongue height ($p = 0.0279$). These findings support the view that there are language-specific differences in default vocal tract settings, and are consistent with the prediction of Kelso et al. [1985] and Vatikiotis-Bateson's [1988] model that an equilibrium position must be a specified target. It is additionally interesting to note that velum and jaw positions were not significantly different across these languages. While there is a relatively strong tendency for English to have a lower jaw ($p < 0.1$; this tendency is discussed further in the 'General Discussion' section), the velum positions show no difference at all. This is somewhat surprising given the substantial differences in nasalization between English and French (e.g., contrastive nasal vowels in French, etc.). However, a straightforward explanation for this may simply be that inhalation between utterances has a greater physiological effect on velum position than on other articulators, masking any language-specific differences that may otherwise have appeared. From our data it is impossible to ascertain whether or not subjects inhaled consistently between utterances. If they did, this may have had a lowering effect on the velum position. Previous studies suggest that inhalation is unlikely to have affected tongue position as it does not have a significant effect on the anteroposterior width of the pharynx [Schwab et al., 1993], and any effect that it does have is smallest in the retroglossal region [Trudo et al., 1998].

A notable observation from these results is that French speakers had: (1) greater pharynx width (advanced tongue root), (2) greater tongue tip-to-alveolar ridge distance (retracted/lowered tongue tip), (3) greater tongue body-to-palate distance (lowered tongue body), and (4) an almost significantly higher jaw position. The vocal tract thus appears to be more open all along its length during the French speakers' ISPs than during the English speakers', despite the jaw being higher. These factors conspire to give the obviously false impression at first glance that French speakers' tongues are smaller than those of English speakers. A more likely explanation for this difference may be that the French AS has a substantial constriction in a region not measured in this study (e.g., the uvular region). Another possible factor that may underlie this (or any) apparent difference in midsagittal area is lateral expansion/contraction of the tongue, which may be another important parameter to consider in future studies of AS. As tongue movement must be dependent on global constraints such as the fixed volume of the tongue [Gick, 2002b; Kier and Smith, 1985], any lateral displacement must be reflected sagittally. It should also be noted that medial compression of the tongue could play a part in the dorsal constriction appearing in lateral consonants [McDowell, 2004] and the reduced pharynx width in English speakers [Narayanan et al., 1997]. This explanation may further correspond with Honikman's [1964] description of the English tongue tip as being 'tapered', suggesting a possible lateral contraction (and hence a midsagittal expansion) in English. Analogous to this, Takano et al. [2002] found that in Japanese, front vowels are associated with lateral expansion whereas back vowels are associated with medial compression. This type of lateral volume change would not be directly evident from a sagittal X-ray image, but future work using coronal ultrasound images may be able to address this possibility.

Experiment 2

In experiment 2, vocal tract measurements during inter-utterance rest positions were compared with similar measurements taken from the target vowel /i/ to test whether the settings observed in experiment 1 function as targets. Variability was compared across the two types of events to determine whether the accuracy of movements into rest position is similar to that of movements to a specified articulatory target. The vowel /i/ was chosen because it has been shown to display very little variability [Perkell and Nelson, 1985], thus providing the most stringent standard for comparison to rest position.

Methods

The same 10 speakers as described in experiment 1 again served as subjects. Target frames for utterances of the vowel /i/ were identified in each X-ray film using Adobe Premiere. For each token of /i/, the frame chosen for analysis was the middle frame in a series of frames in which tongue body height had reached a maximum. Procedures were identical to those used in experiment 1, with measurements taken in NIH Image, and statistics (ANOVAs) calculated using StatView 5.0.1. Measurement locations were identical to the first five measurements taken for experiment 1. The lip measurements were not used in experiment 2 because the lips are not protruded in the production of /i/. Especially in stressed syllables with /i/, it is possible that there is some retraction due to lip spreading. This is currently being investigated using more accurate Optotrak measurements of the lips.

Normalization as discussed in experiment 1 was applied to all raw measurements prior to any statistical comparisons. For this experiment, statistical analyses were done in two stages. First, the standard deviations of both rest and /i/ positions were calculated for each subject individually. This yielded 10 standard deviations each for rest and /i/ positions at each of the five measurement locations. Velopharyngeal port width was excluded because it had values of 0 for most tokens of /i/ except in predictable nasal environments. Second, because the raw measurements had been normalized, comparison of the standard deviations as a whole was possible, with each token treated as a measurement. This allowed for ANOVAs to be calculated to test for significant differences in standard deviations between rest and /i/ positions.

Results: Rest versus /i/ Positions – Standard Deviations

Standard deviation results for each subject are presented in table 3. Statistical results indicate no significant difference in standard deviations between rest and /i/ conditions (pharynx width, $p = 0.8963$; tongue body-to-palate distance, $p = 0.0551$; tongue tip-to-alveolar ridge distance, $p = 0.6309$, and jaw aperture, $p = 0.6515$).

Since the above analysis combines the two languages being investigated, an additional ANOVA analysis was conducted in which language was a predictor, to confirm that this finding does not depend on just one of the two languages. The results of this split analysis were the same as the combined analysis, with no significant difference

Table 3. Standard deviations by subject and vocal tract measurement for rest and /i/

	Position	Pharynx width	Tongue tip to alveolar ridge	Tongue body to palate	Upper to lower jaw
Subject 1	/i/	1.884	8.546	3.992	2.229
	rest	2.565	3.522	2.207	2.555
Subject 2	/i/	5.311	5.222	3.604	0.654
	rest	3.473	2.679	4.677	2.330
Subject 3	/i/	2.140	3.662	1.230	3.082
	rest	2.787	3.080	1.803	2.881
Subject 4	/i/	2.483	1.752	3.211	4.994
	rest	2.757	3.396	3.098	5.119
Subject 5	/i/	3.390	1.790	1.680	3.793
	rest	3.675	1.518	3.468	2.263
Subject 6	/i/	4.469	4.341	2.253	3.817
	rest	3.958	3.789	5.124	5.185
Subject 7	/i/	4.312	2.646	3.752	4.974
	rest	4.300	2.890	3.054	3.245
Subject 8	/i/	2.772	NA	1.456	2.596
	rest	1.930	NA	2.093	3.327
Subject 9	/i/	4.648	2.887	2.429	4.417
	rest	2.384	4.743	4.778	1.416
Subject 10	/i/	3.554	2.293	0.687	2.582
	rest	6.408	4.070	5.983	2.143

NA = Not available.

observed between rest and /i/ conditions at any measurement location ($p > 0.05$ for all comparisons for each language).

Discussion of Experiment 2

Experiment 2 tested a hypothesis about the nature of the language-specific ASs observed in experiment 1. If, as the equilibrium point hypothesis [Perrier et al., 1996] proposes, equilibrium positions are dictated by specified threshold muscle lengths under some form of neural command, then ISPs should be attained with an accuracy similar to that of actual speech target positions.

Results indicate no significant differences in standard deviations between rest position and /i/ (table 3). Thus, the ISP does not appear to be a transition point solely determined by immediately surrounding sounds, since the position is as tightly specified as actual speech targets. These findings support the hypothesis that the ISP is specified as a speech target, in accord with the predictions of the equilibrium point hypothesis and other production models discussed in the 'Introduction' above. This position can certainly be influenced by surrounding events, just as coarticulation affects the physical realization of other speech targets, but it nevertheless behaves, at least in terms of spatial accuracy, like a language-specific, specified target. This further suggests that a language's ISP may be specified in some way in the grammar of a language, possibly as part of the phonetic or phonological inventory. In other words, if the ISP for a language is indeed ultimately determined to be a specified target, then this target must be acquired and stored as part of the information associated with that language.

Because we are arguing from negative evidence in experiment 2, it is perhaps a reason for concern that the statistical results for tongue body-to-palate distance were almost significant. It is possible that the number of tokens in our analysis was too small to find a difference in this measurement between rest position and /i/. A current follow-up study to this one using ultrasound and Optotrak is being conducted with a greater number of tokens and subjects.

If ISPs are specified targets, the question remains as to where these targets come from. As suggested as early as Wilkins [1668], the default vocal tract configuration may fall out as an effect of the frequency of individual speech sounds in the inventory of a given language. This question is also being explored in ongoing studies. An additional issue to consider in future work is whether or not some sections of the vocal tract may be in their target rest positions while other sections are still achieving targets. Further testing of ISPs as specified targets could easily come from studies of kinematic velocity profiles of transitional movements into ISPs.

General Discussion and Further Considerations

The experiments described above show that there does exist a language-specific posture to which the articulators return between utterances, and that this posture is attained with an accuracy comparable to that of specified speech targets. The task remains to compare these ISPs with previously described ASs that occur during speech to determine whether these ISPs are merely default 'recoil' positions active only during pauses between speech, or whether they in fact exercise an active influence on speech. As discussed above, there have been many previous observations from descriptive studies of speech that have implicated certain underlying articulatory differences between English and French. These observations include both detailed descriptions of broad, language-specific tendencies as well as the physical characteristics of English and French 'neutral' vowels (e.g., schwa and filled pauses).

Comparison with Previous Descriptions

Honikman [1964] and Sweet [1890] both described the AS for English and French, and although it is enlightening to analyze their impressions, it should be noted that they are referring to Received Pronunciation (RP) English and European French, which could certainly have a somewhat different AS from the Canadian English and French found in the X-ray data we analyze. Honikman [1964, p. 81] observed that whereas in English the tongue body is anchored 'to [the] roof laterally', in French it is anchored 'to [the] floor centrally', and whereas the tongue tip is 'tapered' in English, it is 'untapered' in French. Because the present experiment relies on X-ray data, we can only discuss indirectly off-midline observations such as these. However, recall from the results of experiment 1 that in French both the tongue body and tongue tip are significantly lower than in English. This agrees with Honikman's [1964] statement that in French the tongue is anchored to the floor of the mouth. Further, if the English AS has a tapered tongue tip, given that the tongue is volume-preserving, we may expect this narrowing of the tongue anterior (depending on the length and degree of the narrowing) to cause a greater constriction towards the upper or rear wall of the vocal tract (i.e., tongue body or tongue tip moving toward the palate, or tongue root retracting into the pharynx). The present results agree with all three of these scenarios for tongue position: English has a narrower pharynx width, a higher tongue body, and a higher tongue tip. As for lip position, Honikman [1964, p. 81] observes that in English the lips are in neutral position and are 'moderately active', while in French the lips are rounded and are 'vigorously active in spreading and rounding'. Assuming that 'rounding' here can be interpreted as lip protrusion, Honikman's [1964] description corresponds with the lower lip findings of experiment 1, but does not agree with our findings for the upper lip. In partial contrast, Sweet [1890, p. 72] suggests impressionistically that overall in English the tongue is lowered, flattened and drawn back from the teeth, and the lips are in a neutral position, while in French the tongue is arched, raised, and advanced, and the lips 'articulate with energy'. Here, although Sweet's [1890] description of the lips can be interpreted as concurring with both Honikman's [1964] description and to some extent with the present findings, his description of the position of the anterior part of the tongue differs in nearly every respect from both of these. It may be that impressionistic description of the position of the tongue, which is out of view during speech, is less reliable than direct measurement using the modern tools of medical imaging. However, this does not explain why Honikman's [1964] impressionistic description agrees more closely with our measurements than Sweet's [1890]. More likely, it may simply be that the apparent conflict is based on differences between the dialects being observed, given the regional differences and the many years of potential language change intervening between the three studies. In any case, it is clear that at least one previous description of language-specific ASs corresponds closely with the findings of the present study.

One incidental finding, not presented above, of this study shows that there are significant differences in rest position between males and females (combined across both languages), with males having significantly greater pharynx width ($p < 0.0001$), and tongue tip-to-alveolar ridge distance ($p = 0.0011$), and females having a greater upper-to-lower jaw distance ($p = 0.0034$). These findings may presumably be accounted for by sex-related differences in vocal tract morphology. For example, Fant [1973] and Fitch and Giedd [1999] found that adult males have a disproportionately longer pharynx in comparison with adult females. Rosas et al. [2002] also found, in addition to size-related differences between men and women, independent sexual dimorphism in mandible size and shape; they attributed these gender-specific differences to musculoskeletal development and growth trajectories. It is thus an important methodological note for future studies of AS that gender must be carefully controlled across language groups – as was serendipitously the case in the database used for the present study, with 2 males and 3 females in each language group.

Comparison with Neutral Vowels

As discussed in the 'Introduction' section of this paper, it may be that a language's neutral 'schwa' vowel is produced with the articulators in a configuration that is closest to that language's underlying AS. Story and Titze [2002] found that the shape of the vocal tract affects the quality of the voice, especially of the neutral vowels. Thus, if there is an underlying default articulatory configuration specific to a given language, it could explain variations in the phonetic characteristics of schwa across different languages. In support of this view is the observation that schwa in English and French are produced differently. Price [1991, p. 77] points out that French schwa is 'pronounced rather further forward in the mouth and with noticeably rounded lips' than English schwa, while Gick [2002a] finds that American English schwa is produced with a retracted tongue root. These differences in English and French schwa correspond closely with differences in ISP observed in this paper, where the tongue root is more retracted (i.e. the pharynx width is smaller) in English than in French, and the lower lip is more protruded in French than in English. However, it is also clear from previous work on English that schwa is not simply a vocalized instance of the ISP. Gick [2002a] shows that the tongue root is not only retracted, but that it retracts beyond the ISP position, i.e., it must have an actively retracted tongue root as part of its target. This finding is supported by Palethorpe and Cox [2003]. The present findings thus suggest that the language-specific characteristics of schwa may be related at least in part to ASs, but that schwa is not necessarily simply a voiced version of a language's AS (at least not in English). That said, 1 subject in Gick's [2002a] study of English showed a bimodal distinction between schwa in lexical words, which showed the active tongue retraction, and schwa in function words, which showed no difference from speech rest position. Thus, it may be that at least some types of schwa bear no target of their own, and reflect directly the AS posture. Continuing work is testing these possibilities.

In addition to schwa, another possible 'neutral vowel' context relevant to AS may be the so-called 'filled pauses' like *uh* or *um* in English. One hypothesis is that phonation during the production of an underlying AS will result in a filled pause-type speech output. Like schwa, however, these filled pauses apparently have targets of their own. For example, Shriberg and Lickley [1993] and Shriberg [1994] find that filled pauses have a lower F_0 than surrounding speech sounds, while Clark [2002] and Clark and Fox Tree [2002] found that filled pauses differ depending on their function: They interact with prosody, have conventional phonological form and independent semantic meaning (discourse functions). These authors thus conclude that *uh* and *um* are in fact phonological words. Thus, as with schwa, while it cannot be the case that filled pauses are simply vocalizations of a language's AS, it is nevertheless possible that language-specific ASs may have a bearing on the realization of these neutral segments.

Conclusions

The findings in this paper support a view in which ISPs assumed between speech utterances: (a) are language-specific; (b) function as active targets; (c) are active during speech, corresponding with the notion of ASs, and (d) exert measurable influences on speech targets, most notably including effects on the properties of neutral vowels such as schwa. The existence of these language-specific ASs further implies far-reaching effects on languages' phonetic and phonological systems, including synchronic and diachronic interactions with the segmental inventory (presumably in both directions) and with phonological processes (assimilation, harmony, etc.). These ASs must also function in language acquisition, sociolinguistic factors, and so on.

Finally, the finding that the ISP represents a language-specific target configuration has important implications for foreign-language teaching. The results give much-needed quantitative evidence to substantiate AS-based methodologies such as those proposed by Mompeán-González [2003] for RP English, and by Collins and Mees [1995] and Esling and Wong [1983] for General American English. Our results indicating a relatively high tongue body in GA English are consistent with Esling and Wong's [1983, p. 91] description of a 'palatalized tongue body position', and our results indicating a relatively high tongue tip in GA English are consistent with Collins and Mees' [1995, p. 418] description of an 'alveolarized' tongue setting.

The present study simply confirms experimentally a phenomenon familiar to generations of previous students of language. However, like some other aspects of the speech signal, such as intonation, AS has long remained beyond the pale for many researchers. The present paper will have achieved its purpose if it can help to bring AS into the realm of rigorous consideration, at least by suggesting methods for direct measurement, and by linking together previous work on the subject. In any case, many future studies will be needed to address the questions remaining about the nature and origins of AS, and influence on speech.

Acknowledgments

Thanks to Sonya Bird, Jason Brown, Fiona Campbell, Monique Charest, Elizabeth Johnson, Elliot Moreton, Shaffiq Rahemtulla, Mark Tatham, Eric Vatikiotis-Bateson, and attendees at the 2003 Nashville meeting of the Acoustical Society of America for helpful comments. This work was funded by an NSERC Discovery Grant to the first author and by NIH grant DC-02717 to Haskins Laboratories.

References

Barry, W.: Comments on chapter 2 (Browman and Goldstein); in Docherty, Ladd, Papers in laboratory phonology II: gesture, segment, prosody, pp. 65–67 (Cambridge University Press, Cambridge 1992).

Bradlow, A. R.: A comparative acoustic study of English and Spanish vowels. J. acoust. Soc. Am. *97:* 1916–1924 (1995).

Browman, C.P.; Goldstein, L.: 'Targetless' schwa: an articulatory analysis; in Docherty, Ladd, Papers in laboratory phonology II: gesture, segment, prosody, pp. 26–56 (Cambridge University Press, Cambridge 1992).

Burris, B.G.; Harris, E.F.: Identification of race and sex from palate dimensions. J. forens. Sci. *43:* 959–963 (1998).

Byrne, D.; Dillon, H.; Tran, K.; Arlinger, S.; Wilbraham, K.; Cox, R.; Hagerman, B.; Hetu, R.; Kei, J.; Lui, C.; Kiessling, J.; Kotby, M.N.; Nasser, N.H.A.; El Kholy, W.A.H.; Nakanishi, Y.; Oyer, H.; Powell, R.; Stephens, D.; Meredith, R.; Sirimanna, T.; Tavartkiladze, G.; Frolenkov, G.I.; Westerman, S.; Ludvigsen, C.: An international comparison of long-term average speech spectra. J. acoust. Soc. Am. *96:* 2108–2120 (1994).

Clark, H.H.: Speaking in time. Speech Commun. *36:* 5–13 (2002).

Clark, H.H.; Fox Tree, J.E.: Using *uh* and *um* in spontaneous speaking. Cognition *84:* 73–111 (2002).

Collins, B.; Mees, I.M.: Approaches to articulatory setting in foreign-language teaching; in Lewis, Studies in general and English phonetics: essays in honour of Professor J. D. O'Connor, pp. 415–424 (Routledge, London 1995).

Disner, S. F.: Evaluation of vowel normalization procedures. J. acoust. Soc. Am. *67:* 253–261 (1980).

Esling, J.H.; Wong, R.F.: Voice quality settings and the teaching of pronunciation. TESOL Q. *17:* 89–95 (1983).

Fant, G.: Speech sounds and features (MIT Press, Cambridge 1973).

Fitch, W.T.; Giedd, J.: Morphology and development of the human vocal tract: a study using magnetic resonance imaging. J. acoust. Soc. Am. *106:* 1511–1522 (1999).

Gick, B.: An X-ray investigation of pharyngeal constriction in American English schwa. Phonetica *59:* 38–48 (2002a).

Gick, B.: Exercises in absurdity: interactions between flapping and tongue position in American English with implications for speech motor planning. J. acoust. Soc. Am. *112:* 2416 (2002b).

Harmegnies, B.; Landercy, A.: Language features in the long-term average spectrum. Rev. Phonét. appl. *73–75:* 69–80 (1985).

Heffner, R.-M.S.: General phonetics (University of Wisconsin Press, Madison 1950).

Honikman, B.: Articulatory settings; in Abercrombie, Fry, MacCarthy, Scott, Trim, In honour of Daniel Jones, pp. 73–84 (Longman, London 1964).

Jenner, B.: 'Articulatory setting': genealogies of an idea. Historiographia Linguistica *XXVIII:* 121–141 (2001).

Kelso, J.A.S.; Vatikiotis-Bateson, E.; Saltzman, E.L.; Kay, B.: A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modelling. J. acoust. Soc. Am. *77:* 266–280 (1985).

Kier, W.M.; Smith, K.K.: Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. Zool. J. Linn. Soc. *83:* 307–324 (1985).

Laver, J.: The concept of articulatory settings: an historical survey. Historiographia Linguistica *V:* 1–14 (1978).

McDowell, R.: Retraction in Montana Salish lateral consonants: an ultrasonic study; MA thesis University of British Columbia (2004).

Mompeán-González, J.A.: Pedagogical tools for teaching articulatory setting. ICPhS 15, Barcelona 2003.
Munhall, K.G.; Vatikiotis-Bateson, E.; Tohkura, Y.: X-ray film database for speech research, ATR technical report TR-H-116 (ATR Human Information Processing Research Laboratories, Kyoto 1994).
Narayanan, S.; Alwan, A.; Haker, K.: Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I: The laterals. J. acoust. Soc. Am. *101:* 1064–1077 (1997).
O'Connor, J.D.: Phonetics (Penguin, Harmondsworth 1973).
Öhman, S.E.G.: Peripheral motor commands in labial articulation. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., QPSR 4 (1967).
Palethorpe, S.; Cox, F.: Vowel modification in pre-lateral environments. 6th Int. Semin. on Speech Production, Sydney 2003.
Parry, J.J.; Burnett, I.S.; Chicharo, J.F.: Language-specific phonetic structure and the quantisation of the spectral envelope of speech. Speech Commun. *32:* 229–250 (2000).
Perkell, J.S.: Physiology of speech production: results and implications of a quantitative cineradiographic study (MIT Press, Cambridge 1969).
Perkell, J.S.; Nelson, W.L.: Variability in production of the vowels /i/ and /a/. J. acoust. Soc. Am. *77:* 1889–1895 (1985).
Perrier, P.; Ostry, D.J.; Laboissiere, R.: The equilibrium point hypothesis and its application to speech motor control. J. Speech Hear. Res. *39:* 365–379 (1996).
Price, G.: An introduction to French pronunciation (Blackwell, Oxford 1991).
Rochette, C.E.: Les groupes de consonnes en français: étude de l'enchainement articulatoire à l'aide de la radiocinématographie et de l'oscillographie (Les Presses de l'Université Laval, Québec 1973).
Rosas, A.; Bastir, M.; Martinez-Maza, C.; Bermudez de Castro, J.M.: Sexual dimorphism in the Atapuerca-SH hominids: the evidence from the mandibles. J. hum. Evol. *42:* 451–474 (2002).
Rosenbaum, D.A.; Loukopoulos, L.D.; Meulenbroek, R.G.J.; Vaughan, J.; Engelbrecht, S.E.: Planning reaches by evaluating stored postures. Psychol. Rev. *102:* 28–67 (1995).
Rosenbaum, D.A.; Meulenbroek, R.J.; Vaughan, J.; Jansen, C.: Posture-based motion planning: applications to grasping. Psychol. Rev. *108:* 709–734 (2001).
Schwab, R.J.; Gefter, W.B.; Pack, A.I.; Hoffman, E.A.: Dynamic imaging of the upper airway during respiration in normal subjects. J. appl. Physiol. *74:* 1504–1514 (1993).
Shriberg, E.E.: Preliminaries to a theory of speech disfluencies; PhD diss. UC Berkeley (1994).
Shriberg, E.E.; Lickley, R.J.: Intonation of clause-internal filled pauses. Phonetica *50:* 172–179 (1993).
Story, B.; Titze, I.R.: A preliminary study of voice quality transformation based on modifications to the neutral vocal tract area function. J. Phonet. *30:* 485–509 (2002).
Sweet, H.: A primer of phonetics (Clarendon Press, Oxford 1890).
Takano, S.; Honda, K.; Dang, J.: Geometrical analysis of the tongue muscles based on MRI and functional modeling of the tongue. J. acoust. Soc. Am. *112:* 2418 (2002).
Trudo, F.J.; Gefter, W.B.; Welch, K.C.; Gupta, K.B.; Maislin, G.; Schwab, R.J.: State-related changes in upper airway caliber and surrounding soft-tissue structures in normal subjects. Am. J. resp. crit. Care Med. *158:* 1259–1270 (1998).
Vatikiotis-Bateson, E.: Linguistic structure and articulatory dynamics: a cross-language study (Indiana University Linguistics Club, Bloomington 1988). (Publication of author's 1987 PhD diss.)
Wilkins, J.: An essay towards a real character, and a philosophical language (J. Martyn, London 1668/1967). (Reprinted by Scholar Press, Menston 1967.)